Automated Open Circuit Scuba Diver Detection with Low Cost Passive Sonar and Machine Learning

by

Lieutenant Commander Andrew M. Cole, United States Navy

B.S., United States Naval Academy, 2006

Submitted to the Joint Program in Applied Ocean Science & in partial fulfillment of the requirements for the degree of

Master of Science in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION

June 2019

Author . . . . . .

Joint Program in Applied Ocean Science & Engineering

Massachusetts Institute of Technology & Woods Hole Oceanographic Institution

May 10, 2019

Certified by . . . . . .

Carl L. Kaiser

Program Manager

Woods Hole Oceanographic Institution

Thesis Supervisor

Certified by . . . . . .

Andone C. Lavery

Senior Scientist

Woods Hole Oceanographic Institution

Thesis Supervisor

Certified by . . . . . .

Henrik Schmidt

Professor

Massachusetts Institute of Technology

Mechanical Engineering Faculty Reader

Accepted by . . . . . .

Nicolas Hadjiconstantinou

Chair, Mechanical Engineering Committee for Graduate Students

Massachusetts Institute of Technology

Accepted by . . . . . .

David Ralston

Chair, Joint Committee for Applied Ocean Science and Engineering

Massachusetts Institute of Technology

Woods Hole Oceanographic Institution

Automated Open Circuit Diver Detection with Low Cost Passive Sonar and Machine Learning

by

Lieutenant Commander Andrew M. Cole, USN

Submitted to the Joint Program in Applied Ocean Science & Engineering
Massachusetts Institute of Technology
& Woods Hole Oceanographic Institution
on May 10, 2019, in partial fulfillment of the
requirements for the degree of
Master of Science in Mechanical Engineering

Abstract

This thesis evaluates automated open-circuit scuba diver detection using low-cost passive sonar and machine learning. Previous automated passive sonar scuba diver detection systems required matching the frequency of diver breathing transients to that of an assumed diver breathing frequency. Earlier work required prior knowledge of both the number of divers and their breathing rate. Here an image processing approach is used for automated diver detection by implementing a deep convolutional neural network. Image processing was chosen because it is a proven method for sonar classification by trained human operators. The system described here is able to detect a scuba diver from a single acoustic emission from the diver. Twenty dives were conducted in support of this work at the WHOI pier from October 2018 to February 2019. The system, when compared to a trained human operator, correctly classified approximately 93% of the data. When sequential processing techniques were applied, system accuracy rose to 97%. This demonstrated that a combination of low-cost, passive sonar and a properly tuned convolutional neural network can detect divers in a noisy environment to a range of at least 12.49 m (50 feet).

Thesis Supervisor: Carl Kaiser
Title: Program Manager
Woods Hole Oceanographic Institution

Thesis Supervisor: Andone Lavery
Title: Senior Scientist
Woods Hole Oceanographic Institution

Acknowledgments

This research was funded by the U.S. Navy's Civilian Institution Program with the MIT/WHOI Joint Program. The Woods Hole Oceanographic Institution provided resources for the scuba diving conducted during the course of this thesis. I am incredibly grateful to the U.S. Navy Submarine Force for providing the opportunity to study at these two world class institutions and work closely with very talented people.

Thank you to Dr. Carl Kaiser, my primary research advisor, for his mentorship and guidance throughout my time in the Joint Program. With his busy and demanding schedule, he made my research and development a top priority. Additionally Dr. Kaiser ensured that my professional development included more than my research topic. This included visiting several Navy facilities, gaining a better understanding of the synergy between academia and the department of defense, and conducting experiments with autonomous undersea vehicles. He also enabled me to sail on a research ship in support of the AUV Sentry team. This was a unique opportunity that I will never forget.

Thank you to Mr. Edward O'Brien, the WHOI Dive Operations Manager. Without his assistance and support this research would not have been possible. Mr. O'Brien spent countless hours both improving my diving skills and conducting this experiment. The bulk of this research was conducted during winter in Massachusetts with near freezing water and limited visibility. Mr. O'Brien braved these conditions with me to conduct my experiment. I am truly grateful for his support.

Thank you to Dr. Andone Lavery, my co-advisor, for her support and guidance during my time in the MIT/WHOI Joint Program. Her direction was instrumental at reorienting me to academia and ensuring I was set up for success. Dr. Lavery's devotion to the success of all students in the Applied Ocean Sciences and Engineering department of the MIT/WHOI Joint program and her service as the Applied Ocean Physics department education coordinator are very commendable.

Thank you to Professor Henrik Schmidt, my academic advisor, for his support and guidance during my time in the MIT/WHOI Joint Program. Professor Schmidt's counsel enabled my success at MIT. He goes out of his way to help all Navy students in the MIT/WHOI Joint program and much of our success is owed to him. Additionally his instruction in autonomous vehicles and acoustics were both enjoyable and enlightening.

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

Chapter 1

Introduction

1.1 Motivation

Remote detection of scuba divers is important for several reasons include monitoring the frequency of divers at recreational dive sites, identifying the presence of divers at locations where diving is forbidden (e.g. environmentally protected areas, cultural heritage sites such as shipwrecks or underwater archaeology), or preventing diver interference with aquaculture or other critical infrastructure [18][45][31]. To date, diver detection has been conducted primarily using sonar as opposed to other generally shorter range methods such as optical systems or monitoring the surface of the water for the presence of bubbles.

Either active or passive sonar can be used for scuba diver detection. Active sonar can detect divers under a wide range of conditions and has a better chance of detecting closed-circuit divers [13]. The down side to active sonar is that it requires substantial power, is expensive, and has the potential to disturb marine life [31]. Less research has been conducted on diver detection using passive sonar; however, passive sonar has the potential to be low cost, require minimal power, and it does not disturb the surrounding environment [18].

There are two main categories of divers, open-circuit and closed-circuit (rebreather) scuba divers. The need to detect closed-circuit divers is largely a military problem with a high penalty for missed detections. Active sonar is likely a better choice for closed-circuit diver detection, as the noise emitted by closed-circuit divers is generally low and cost and power are less likely of concern, as most closed circuit diver detection is military. Open-circuit scuba divers are more common than closed-circuit scuba divers for civilian diving applications and generally have a louder presence in the water making them more suitable

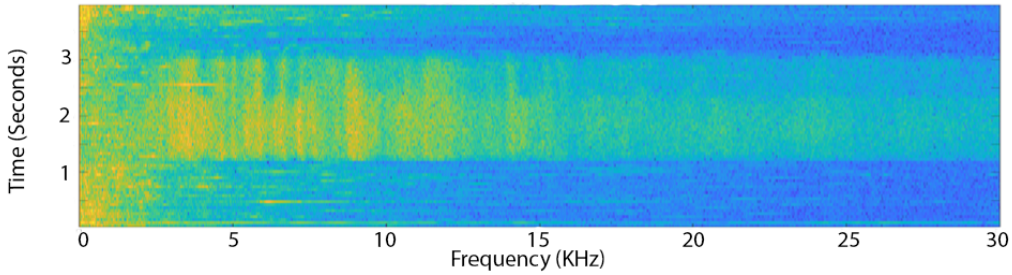


Figure 1-1: Four Second Spectrogram Containing a Diver Transient.

for passive sonar detection. This thesis focuses on the passive acoustic detection of open-circuit scuba divers. Specifically, this thesis will show that machine learning can be used along with low-cost passive acoustic systems to detect divers even, in a noisy environment.

The acoustic signature of divers is neither true narrowband, nor true broadband, making diver detection via passive sonar a non-trivial problem. If the diver signature was defined by a few discrete frequencies, detection would be accomplished using traditional automated means of identifying the presence of those frequencies. Conversely, if the diver signature was only broadband it would be indistinguishable from other broadband transients of the same duration. In this case, detection would require the method explored and is discussed in section 2.1. Instead, the diver signature is broadband with varying amplitudes across the frequency spectrum. The amplitudes and frequency bands are dependent both on the diver's equipment and operating depth [15][33]. Figure 1-1 shows a four second spectrogram containing a diver transient. Frequency is on the X axis and time is on the Y axis. Here machine learning algorithms are used to automate proven sonar interpretation techniques for such problems. The ability to automatically detect this type of signature using passive sonar likely has other uses that are outside the scope of this thesis.

Diver detection with passive sonar may have limitations when conducted in a real world port environment. This work is discussed in section 2.1. The common thread in previous work is it attempted to use traditional methods for diver detection and experiments were frequently conducted in controlled environments that were not representative of the real world ocean environment where detection is most needed. Existing techniques require long integration times and are highly susceptible to interference by broadband transients.

This thesis explored a different method of automated diver detection. Instead of using traditional methods of automated diver detection, discussed in chapter 2; machine learning

based image processing techniques were used. This method has been demonstrated to be effective for ship classification by the U.S. Navy and the author has experience with this approach [14][8][44][16]. This thesis extends this approach by replacing a trained human sonar operator with a machine learning model and classifying scuba divers as opposed to ships.

1.2 Thesis Overview

This thesis evaluates the feasibility of using low-cost passive sonar and an image processing based machine learning technique for automated open-circuit scuba diver detection. Image processing techniques had been previously used for sonar classification, but this thesis extends this approach by replacing a trained human operator with a machine learning algorithm. Chapter 2 presents previous work in the two key fields addressed by this research, automated open-circuit diver detection with passive sonar and image classification using deep convolutional neural networks.

The methods used in this research are addressed in chapter 3. It starts off by discussing the diving protocol for the experiment. The chapter then moves into signal processing and a discussion of manually evaluation and labeling of the acoustic data. Machine learning is then covered by outlining the selection of metrics, software, and the specific network chosen. Chapter 3 next walks through the process of optimizing the machine learning model and an alternative method for acoustic data evaluation. The chapter is concluded by examining alternate methods to diver detection and acoustic modeling pertinent to diver detection.

Chapter 4 presents the results of this thesis. It examines the machine learning model performance with both validation data and data collected after the model generation was complete. The machine learning model's ability to detect divers during periods of high background noise and the results of processing data in a sequential manner are then discussed. The chapter concludes by evaluating the acoustic propagation paths between the diver and the hydrophone.

The overall conclusions of this thesis and recommended follow on work are discussed in chapter 5. This chapter notes multiple discoveries and adds independent confirmation to conclusions reached by other researchers. Chapter 5 discusses the original contribution of this thesis; a combination of low cost, passive sonar and a properly tuned convolutional

neural network can detect divers in a noisy environment to a range of at least 12.49 m (50 feet).

Figure 1-2: Construction at Martha's Vineyard Ferry Terminal, 18 January 2019.

Chapter 2

Background

This thesis evaluates the feasibility of using low-cost, passive sonar and machine learning to conduct automated detection of scuba divers. There has been significant work conducted separately for both automated diver detection using passive sonar and object classification using machine learning. However, the author was only able to identify a single prior attempt at combining the two [51]. The earlier use of machine learning for diver detection followed the same workflow as other diver detection work, but used a support vector machine (SVM) scheme for classification. No previous attempts at using an image processing approach for diver detection were identified. This thesis uses an image processing approach via a deep convolutional neural network for diver detection.

This chapter presents relevant previous work in both automated diver detection using passive sonar and image classification using machine learning. The first section of this chapter focuses on earlier diver detection work while the second section looks at machine learning approaches. Although the literature on machine learning is extensive, this chapter will progressively focus on specific aspects of machine learning relevant to the methods used here.

2.1 Diver Detection Literature Review

Significant previous work exists in automated detection of open circuit scuba divers using passive sonar. Most reported research was conducted in pools or isolated bodies of water where interfering contacts and background noise were not of concern, and therefore does not account for the actual conditions real-world port environments. Additionally, most prior

work required long integration times for detection, making detection difficult in low signal to noise ratio (SNR) environments or where the signature is intermittently observable.

Diver detection with passive sonar requires the diver to emit noise from either their movement or equipment. Since the human body is relatively quiet, the acoustic emissions are likely to come principally from equipment. Divers wear wetsuits or dry-suits to regulate their body temperature, fins to assist in swimming, masks to see effectively underwater, and lead weights and a buoyancy compensation device to achieve neutral buoyancy. Divers also wear equipment to breathe under water consisting of a tank and series of regulators. The tank is filled with a compressed gas capable of sustaining human life. A first stage pressure regulator is attached directly to the tank. When the diver inhales it reduces the compressed gas pressure from approximately 20,700 kPa (3,000 psia) for a full tank to approximately 1,030 kPa (150 psia). The second stage pressure regulator, which is the piece of equipment placed in the diver's mouth, reduces the gas pressure further to near ambient sea pressure, allowing the diver to breathe underwater [15].

Donskoy *et. al.* studied the acoustic emissions of open circuit scuba divers to identify their sources and characteristics [15]. They conducted a series of experiments where only specific parts of the dive equipment, such as the tank, the first stage regulator, or the second stage regulator were submerged in a tank with hydrophones. The rest of the equipment was in the air, acoustically isolating it from the hydrophone. Donskoy *et. al.* also conducted experiments where the diver was submerged in a pool with full equipment. Through these experiments they identified the primary acoustic source for open circuit scuba divers was the first stage regulator [15]. Donskoy *et. al.* also established that the diver's acoustic signature depended both on the equipment and the diver. The age and model of the first stage regulator were the dominate factors; however, they also showed that the pressure in the tank and the diver's experience and activity also played a role in the acoustic signature [15]. They did not evaluate the effect of depth or water temperature on the diver's acoustic characteristics, though it is possible that both of these factors also play a role.

Lohrasbipeydeh *et. al.* confirmed Donskoy's conclusion that the dominate diver acoustic emission is diver inhalation [31]. Lohrasbipeydeh noted that the diver inhalation transient is 40 dB louder than the exhalation transient and driven primarily by the first stage regulator. He suggested that this is likely due to the higher differential pressure across the first stage regulator, with a differential pressure of 8-190 bar, in contrast to the 0.4 bar across the

second stage regulator.

All previous work identified by the author relied on detecting periodic transients that met a pre-determined threshold for possible diver transients. If the transients occurred at a frequency that fell in a pre-determined diver breathing frequency band, the series of detections were classified as a diver. The frequency ranges used by various researchers are presented in table 2.1. If the period of the transients fell outside of this band, diver detection was not indicated. The only major difference between studies was the signal processing technique used to extract the transients and evaluate their periodicity. Although the signal processing approaches were substantially different, all investigators approached automated diver detection based on periodicity of transients. For the remainder of this thesis periodicity of transient based diver detection will be referred to as traditional automated diver detection. Tables 2.1 - 2.4 show various parameters used and results of prior research.

For traditional automated diver detection to succeed the environment needs to be free of non-diver periodic transients that meet the pre-determined criteria for a diver transient. This assumption is generally good in a controlled environment such as a pool or pond but may not be valid in the presence of normal marine background noise. Johansson *et. al.* noted that in an environment where short duration transients are present, such as locations near construction, there is a lower probability of detection and an increased likelihood of false alarms [23]. It is possible to tune the detection system to avoid a specific transient type; however, this requires a non-trivial amount of effort and foreknowledge of the transients that will be encountered.

Another requirement for traditional diver detection is the detection of every sequential transient over the integration time. The detection time required for the algorithm proposed by Sharma *et. al.* was 240 seconds, or four minutes [42]. The only other paper identified by the author that specified the required integration time for their system was Sun *et. al.*, which suggested that 10 breathing cycles, or approximately 30 seconds were required [46]. Based on the experience of the author, a diver breathing period of 3 seconds is too low. The diver breathing rate presented by Sharma *et. al.* of 5 to 12 seconds is more realistic for an experienced diver, leading to integration times of 50 to 120 seconds [42]. In benign environments detection time is not a problem. However, in real world marine environments there are interfering contacts, such as boats, that can conceal the diver transients [31]. Masking the diver transients during the required integration time will result in a missed

detection.

Another important set of assumptions for traditional automated diver detection is that the number of divers and their breathing frequency is known a-priori. This assumption is the least transferable to real world problems and environments. The majority of the papers reviewed only consider a single diver. The only exception discovered by the author was Lennartsson *et. al.*, who considered two divers [28]. Safety concerns compel most divers to dive in groups of at least two. For this reason, a diver detection system tuned to detect a single diver may fail to detect many real-world divers. It is common for divers to dive in groups of two to more than ten. Even assuming a maximum of four divers, the breathing frequency range grows significantly in size and a-periodicity, potentially confusing period based detections.

Zhao *et. al.* made the first reported use of machine learning for diver detection [51]. Their experiment was conducted in a pool and looked for diver transients in the band of 49 - 51 kHz. Unlike traditional automated diver detection, they trained a SVM machine learning algorithm for final classification [12]. The model they used was trained using 300 10 second samples [51]. Zhao demonstrated that this method outperformed traditional automated diver detection when a diver breaths aperiodically, however this classifier was still based on detecting diver transients at a rate consistent with an assumed diver breathing frequency. As such it was subject to the same limitations as outlined previously.

Another diver detection method was presented by Sattar and Dudek [40]. They conducted diver detection using electro-magnetic cameras mounted to a remotely operated vehicle. Computer vision was applied to the image stream. The intention of their system was to enable an underwater vehicle to detect and follow a scuba diver. Though this idea is innovative, and likely useful for robot-diver coordination, the use of optical cameras is not well suited for more generalized diver detection. Electro-magnetic cameras rely on significant ambient light and good optical clarity, both of which are often lacking in the underwater environment [40].

Table 2.1: Reported Diver Breathing Frequencies [46] [30] [28] [45] [11] [51] [10] [9] [42].

Article	Diver Breathing Frequency	Diver Breathing Period
Sun *et. al.*	0.33 Hz	3 Seconds
Lo *et. al.*	0.15 – 0.5 Hz	2 – 6.7 Seconds
Lennartsson *et. al.*	0.3 – 0.4 Hz	2.5 – 3.3 Seconds
Stolkin *et. al.*	0.2 – 0.4 Hz	2.5 – 5 Seconds
Chung and Li	0.2 – 0.3 Hz	3.3 – 5 Seconds
Zhao *et. al.*	0.3 Hz	3.3 Seconds
Chen *et. al.*	0.3 Hz	3.3 Seconds
Chen and Tureli	0.3 Hz	3.3 Seconds
Sharma *et. al.*	0.083 – 0.2 Hz	5 – 12 Seconds

Table 2.2: Maximum Reported Diver Detection Range [46] [28] [45] [31] [23] [9] [42].

Article	Maximum Passive Diver Detection Range
Sun *et. al.*	70 m
Lennartsson *et. al.*	30 m
Stolkin *et. al.*	18.29 m (60 feet)
Lohrasbipeydeh *et. al.*	10 m
Johansson *et. al.*	18 m
Chen and Tureli	10 m
Sharma *et. al.*	150 m

Table 2.3: Frequency Band for Diver Detection [46] [24] [28] [51] [23] [30] [18].

Article	Frequency Band for Diver Detection
Sun *et. al.*	3 – 10 kHz
Korenbaum *et. al.*	200 - 500 Hz
Lennartsson *et. al.*	30 - 35 kHz
Zhao *et. al.*	49 – 51 kHz
Johansson *et. al.*	0 – 60 kHz
Sun *et. al.*	20 – 60 kHz
Lo *et. al.*	35 – 80 kHz
Hari *et. al.*	25 – 75 kHz

Table 2.4: Diver Detection Testing Locations [23] [10] [9] [42] [15] [46] [18] [28] [45] [51] [33] [11].

Article	Testing Location	Article	Testing Location
Johansson *et. al.*	Gothenburg Harbor, Sweden	Lennartsson *et. al.*	Gothenburg, Sweden
Chen *et. al.*	Hudson River, Hoboken, NJ	Stolkin *et. al.*	Hudson River, Hoboken, NJ
Chen and Tureli	Hudson River, Hoboken, NJ	Molyboha and Zabarankin	Hudson River, Hoboken, NJ
Sharma *et. al.*	Port of Davisville, RI	Lohrasbipeydeh *et. al.*	Sannich Inlet, Victoria, BC
Donskoy *et. al.*	Pool	Chung and Li	Hudson River, Hoboken, NJ
Sun *et. al.*	Songhua Lake, Jiling, China	Zhao *et. al.*	Pool
Hari *et. al.*	Singapore		

2.2 Machine Learning Literature Review

Machine learning is used for a variety of tasks including automated spam email detection, online marketing, electronic media recommendations, self driving cars, speech recognition, text understanding and translation, and a variety of computer vision tasks such as facial recognition and image classification. There are three general categories of machine learning, unsupervised, supervised, and reinforcement learning. In unsupervised machine learning the network is provided data that is not labeled, and draws inferences about the data without human input.

Supervised learning requires a labeled training data set. The training set is fed into the machine learning network and a model is trained to discriminate between the categories of the data. The model is then used to classify non-labeled data.

In reinforcement learning the machine learning model receives feedback after making a classification. The feedback is positive for a correct classification and negative for an incorrect classification. Feedback is used to update the model and improve its performance.

Of the three machine learning categories, supervised learning is currently the most common, but the use of unsupervised learning is expanding [34]. One reason for the growth of unsupervised learning is the high cost required to generate large labeled data sets [34]. This thesis uses supervised learning for automated diver detection which is described in detail in chapter 3. Supervised learning was chosen as it is a relatively mature field compared with the other two options and the cost of labeling data in this application is modest. Unsupervised and reinforcement learning will not be addressed further here . Figure 2-1 is a diagram showing the selection of the machine learning type used in this thesis.

Either an audio or image processing approach would have been suitable for automated open circuit diver detection with passive sonar. An image processing approach was chosen as this is an established method of classification using sonar; however, attempting diver detection using the raw acoustic data would be interesting future work and is discussed in chapter 5. A convolutional neural network was selected over other classification methods as convolutional neural networks are the most frequently used method for image classification with machine learning and appeared suitable for automated diver detection [27].

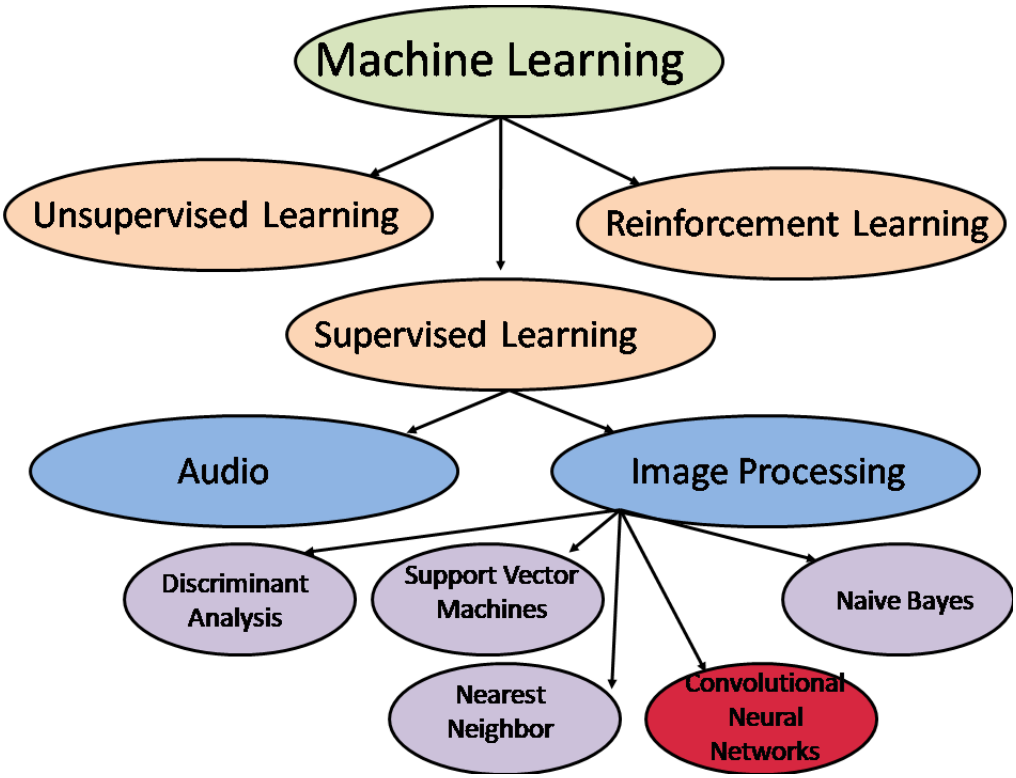


Figure 2-1: Categories of Machine Learning.

2.2.1 Convolutional Neural Network Overview

Deep convolutional neural networks can be used for a variety of tasks and are frequently used for computer vision and speech recognition [27]. Image classification, a subset of computer vision, is placing an image in a pre-defined category. This task is natural for humans but challenging for automated systems. Image classification can be as simple a binary decision, "does the image contain a cat or not", or more complex such as categorizing all things that appear on the side of the road. In the ImageNet competition, an annual image classification competition, one million images are classified into one thousand different categories [34].

Deep convolutional neural networks contain an input layer, an output layer, and multiple hidden layers in between. The neural network is considered deep if there is more than one layer between the input and output layer, which are called hidden layers. Hidden layers are arraigned into groups called modules. Inside the modules, each layer performs a different function [27]. Modules generally consist of a convolutional layer, a pooling layer, and an activation layer. At the end of the network there is at least one fully connected layer. The

final layer in a network is a classification layer that labels the input image [34]. Figure 2-2 depicts a very simple deep convolutional neural network.

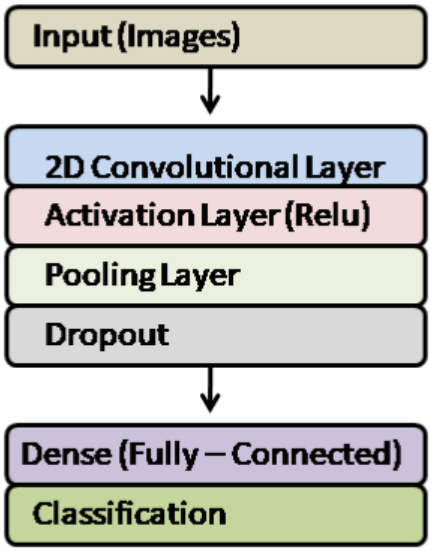


Figure 2-2: Basic Deep Convolutional Neural Network.

Convolutional layers extract features from their input [34]. Each convolutional layer has a receptive field which scans the input. The scan is convolved with learned weights to extract features and create a feature map [27]. An activation layer follows the convolutional layer and performs a manipulation to the feature map, producing an activation map. This enables extraction of non-linear features from the feature map. Pooling layers follow the activation layers and reduce the spatial resolution of the activation map. This results in spatial invariance to translations or distortions of the input [34]. Srivastava *et. al.* showed that dropout, random disconnection of a portion of the neurons in the network, makes a network less susceptible to overfitting, resulting in higher performance [19]. Dropout is common in convolutional neural networks following pooling layers.

At the end of a convolutional neural network there is at least one fully connected layer. In a fully connected layer each neuron is connected to every neuron in the previous layer. Fully connected layers also serve as feature extractors for higher level reasoning [34]. The final layer in a convolutional neural network is a classifier which outputs a label for the input data.

2.2.2 Convolutional Neural Network History

Convolutional neural networks trace their origin to the work of Hubble and Wiesel in 1959 and 1962 [17]. Hubble and Wiesel evaluated the response of cats' brains to visual stimuli to determine the structure and operation of the visual cortex in their brains [21][22]. This work was the foundation of the biological relationship between artificial neural networks and neural networks in the brain of mammals [27]. The first notable use of a neural network based on Hubble and Wiesel's research was by Fukushima in 1974 with a network named Neocognitron [17].

Neocognitron was a neural network consisting of a series of cells connected in a hierarchical manner. The arrangement of the cells was similar to the structure of the visual cortex of cats discovered by Hubble and Wiesel [17]. Though not a convolutional neural network, as it lacked an end to end learning algorithm such as back propagation, this network was the predecessor to modern convolutional neural networks [27]. Neocognitron succeeded at identifying simple input patters, which is a rudimentary form of image classification [17].

The first reported use of convolutional neural networks was in 1989 by Waibel *et. al.* where a time-delayed convolutional neural network was used for speech recognition [27][49]. The first reported use of convolutional neural network for image processing was conducted in the early 1990's in document reading systems. By the end of the 1990's these systems were used for reading more than 10% of the checks written in the United States [43][27]. Between this time and 2012 there was limited reported use of convolutional neural networks. This was likely due largely to the substantial computation power required for convolutional neural networks and the fear that the networks would get stuck in poor local minima of the loss function during training, and therefore would not be optimized correctly [27].

In 2012, convolutional neural networks were accepted by main stream machine learning and computer vision learning communities after a convolutional neural network won the ImageNet competition [27]. LeCun *et. al.* suggest the key developments that enabled a convolutional neural network to win the competition were the use of graphics processing units (GPUs) for computation and rectified linear units (ReLu) as the activation function. The GPUs replaced central processing units (CPUs) for the computation and ReLus replaced sigmoid or hyperbolic tangent functions as the activation function [27]. These changes enabled faster computing with GPUs, conducting training 10-20 times faster than CPUs,

and ReLus conducting training six times faster than hyperbolic tangent activation functions [25][27].

There are emerging trends in the use of convolutional neural networks. Of particular note is the use of extremely deep networks to improve performance. Examples of extremely deep neural networks include GoogLeNet which consisted of 22 layers and won the 2014 ImageNet competition [38] and a network by MSRA which had 152 layers and won the 2015 ImageNet challenge [34]. Additional trends include developing networks that are easily deployed on mobile devices where storage and computation power are limited, as well as dealing with images that contain more than one label [34]. Improving model efficiency with respect to computation has historically been an area of active research and remains so today [34].

In addition to high end machine learning research, there is a trend to make machine learning more accessible through the use of specialized machine learning libraries. The use of dedicated machine learning tools streamlines the programming required for machine learning, allowing the programmer to focus on higher level network structure. The libraries optimized for machine learning include Google TensorFlow, Keras, Caffe, Theano, Scikit-Learn, and Tourch [36][4][1][32].

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Methods

This chapter outlines the methods used in this thesis. The experiment, signal processing, manual data evaluation, machine learning, other methods for diver detection, and physical modeling are each discussed in detail. The results of this analysis are presented in chapter 4.

3.1 Methods Overview

The first step in any supervised machine learning process is to generate a labeled data set. In this case data were collected with scuba divers at specified distances from a low-cost, passive hydrophone seaward of the WHOI dock. The acoustic data from the hydrophone was processed into spectrograms, a time-frequency image format, for human review. The data were then labeled and reformatted for machine learning.

Once an adequate labeled data set was available, a machine learning network was constructed. Spectrograms for machine learning were created, and then split into three groups: training, testing, and validation. The data were feed into a machine learning network for training and tuning. The network produced a model capable of discriminating between data containing diver acoustic emissions and data which only contained background noise. After training, the network was tuned by adjusting parameters and evaluating the model's performance. This process was repeated until an apparent optimum outcome was reached. Figure 3-1 is a flow chart depicting the overall process that is presented in the remainder of this chapter. The optimized model was tested against the validation data which had not been used in the training or tuning processes.

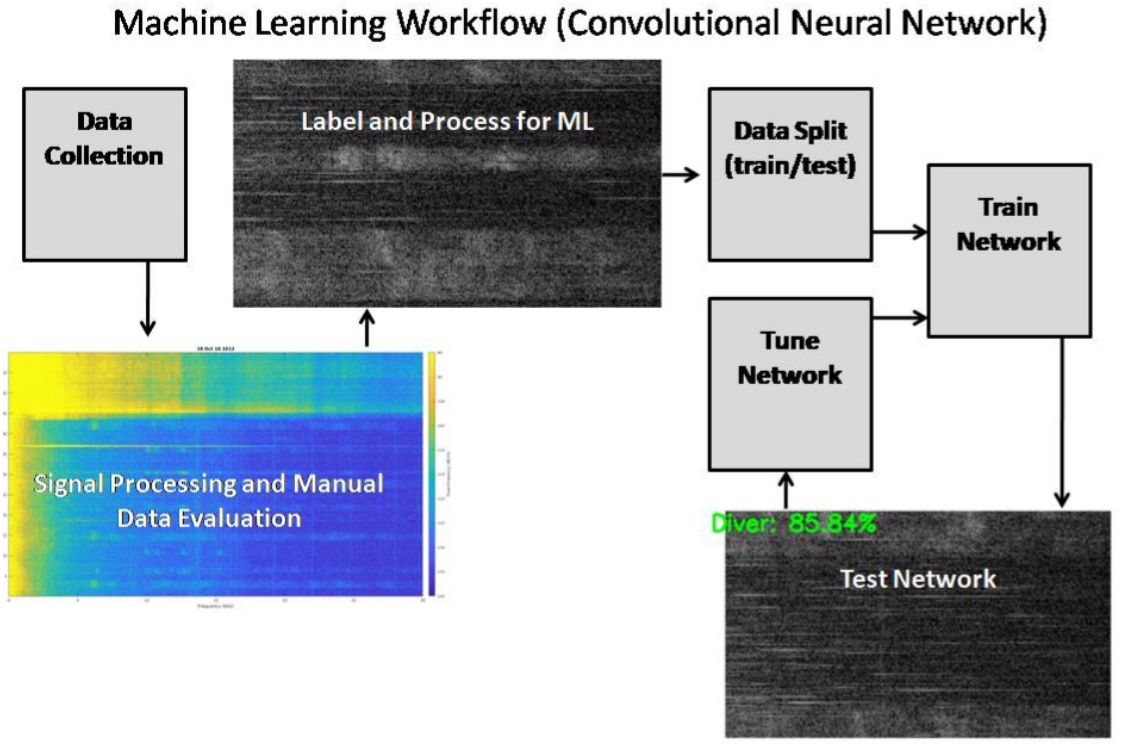


Figure 3-1: Overall Machine Learning Work Flow.

3.2 Experiment Setup and Data Collection

Acoustic data were collected at the WHOI pier when divers were present and when they were not. Divers followed a pre-determined testing protocol to ensure consistent data. Acoustic data were collected from a low-cost passive hydrophone near the WHOI dock and the data were recorded by a National Instruments Data acquisition unit topside.

3.2.1 Testing Site

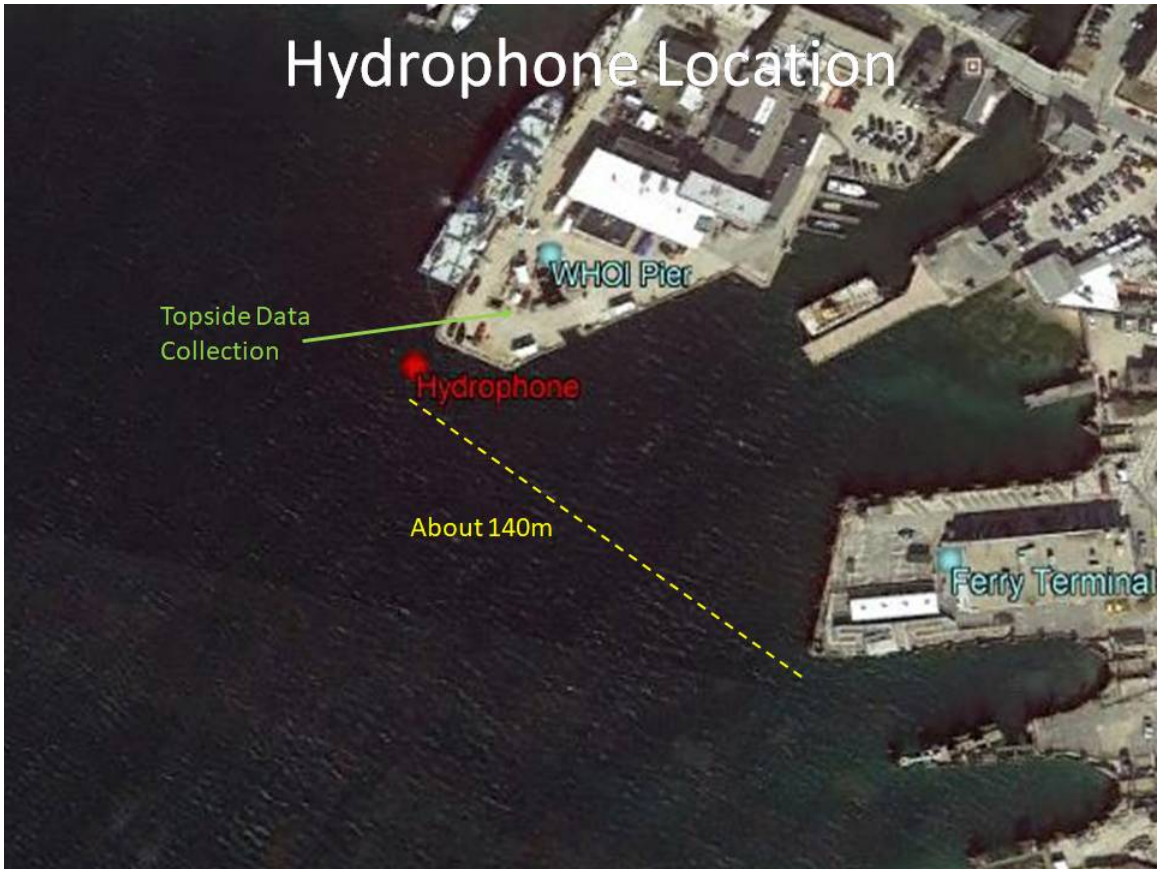


Figure 3-2: Location of Hydrophone used in the Experiment.

The WHOI pier was chosen as the testing location both because it was convenient and because it is representative of a real-world port environment. Confounding factors include interfering surface vessels, noise from research ships moored at the pier, and noise from construction at the nearby Martha's Vineyard ferry terminal. The hydrophone was placed in Vineyard Sound, the body of water between Martha's Vineyard and Woods Hole. It was

approximately 6 m seaward of the southwest corner of the WHOI pier in 21 m of water. In this location it was less 25 m from the primary WHOI berth, which is used by Global and Regional class University-National Oceanographic Laboratory (UNOLS) vessels. The hydrophone was approximately 140 m from the Martha's Vineyard Ferry Terminal and 600 m from the Ram's Island pleasure craft anchorage. It was also directly adjacent to the entrance to the Eel Pond Anchorage. Figure 3-2 is a graphical depiction of the hydrophone's location.

The hydrophone was exposed to noise from vessels in Vineyard Sound, the Martha's Vineyard ferries, construction noise adjacent to the ferry terminal, electrical noise from the WHOI pier, and ventilation and electrical noise from ships moored at the WHOI pier. The background noise at the hydrophone varied significantly from day to day and hour to hour. Vessels underway near the pier frequently masked the acoustic signature of scuba divers. For these reasons it would have been challenging to use traditional methods for diver detection that require a long integration time.

3.2.2 Equipment

The equipment used for data collection was a High Tech Incorporated (HTI) 96 min pre-amplified hydrophone. The pre-amplification improved the signal to noise ratio of the hydrophone's output and minimized the effects of the long cable, which was necessary for this experiment. It had depth rating of 3048 m (10,000 feet), frequency response from 2 Hz to 30 kHz, and a sensitivity of -240 dB. At a length of 6.35 cm and a radius of 1.9 cm the hydrophone was small in size. At a cost of $300 it is affordable for most applications even in quantity. The hydrophone was connected to a National Instruments DAQ series data acquisition unit located on the WHOI pier via a custom pi filter board.

The data were recorded in a Technical Data Management Streaming (TDMS) format which is the standard data format of National Instruments data accusation units. A file duration of one hour was chosen, enabling intuitive data processing and archival. With the selected recording settings, each one hour file was 1.688 MB in size. The recording settings are shown in table 3.1.

3.2.3 Testing Protocol

Data were collected by recording divers at known ranges from the hydrophone. A control data set was created by recording when divers were not in the water. A total of 20 dives

Table 3.1: Hydrophone Recording Settings.

Data Type	TDMS
File Length	1 Hour
File Size	1.688 MB
Recording Frequency	60 kHz
Spectrogram Frequency Range	0-30 kHz

were conducted near the hydrophone. The procedure outlined below was utilized by groups of two to three divers for data collection:

- Enter the water at the WHOI pier instrument well, located approximately 45 m from the hydrophone.
- Descend until approximately one meter above the bottom.
- Swim to the hydrophone, recording the time of arrival.
- Remain within 1.5 m (5 feet) of the hydrophone for 3 to 4 minutes, allowing time for the hydrophone to record the acoustic emissions from the divers.
- Swim to 3.05 m (10 feet) from the hydrophone and remain there for an additional 3 to 4 minutes. Record the time of arrival at this position.
- Repeat the preceding step in 3.05 m (10 foot) intervals, out to a distance of 15.24 m (50 feet), or until forced to conclude the dive due to diver bottom time limits.

On several occasions the procedure above was reversed with divers starting at 15.24 m (50 feet) from the hydrophone and moving closer to it every 3 to 4 minutes. A total of 7 divers, using at least 12 different sets of equipment, conducted the dives for this experiment. Over the course of the 20 dives 5,474 10-second spectrograms were generated; half containing divers and half containing only control data. A copy of the procedure and the data tables that the divers used is contained in appendix B. A list of all dives conducted and associated environmental conditions is in appendix A.

The maximum range of 15.24 m (50 feet) was chosen because it initially appeared to be the limit a human could detect a diver. In later dives, divers were occasionally detectable as soon as they entered the water, approximately 45 m from the hydrophone. A maximum

range of 15.24 m also allowed a diver to complete the entire testing protocol using air as their gas source without exceeding their no-decompression bottom time limit. Limiting range, and thus dive time, was also important during the winter months where dive duration is limited by environmental exposure using standard equipment. Safely extending range and bottom time in the winter would have required highly specialized equipment and training that were not available.

3.3 Signal Processing for Manual Evaluation of Data

Signal processing for this thesis was conducted in MATLAB. Machine learning was conducted as a second step using Python. The acoustic data were displayed in an image based, time-frequency format using the MATLAB spectrogram function. The spectrograms were saved as Portable Network Graphics (PNG) files for manual evaluation. Spectrograms were used because they displayed the frequency content of the signal with respect to time. This is the same format use by the Navy for target classification and is a proven method.

One minute spectrograms were chosen because this duration worked well for evaluation and classification by a human. Using one minute spectrograms divided the one hour long files into manageable chunks and was short enough to identify the characteristics of diver inhalation transients, the loudest acoustic source of open circuit divers [15]. The one minute duration of the spectrograms was only used for manual evaluation and was not related to integration time. Figure 3-3 shows a spectrogram containing emissions from three divers, 9.14 m (30 feet) from the hydrophone. The diver transients are visible for the first (bottom) 42 seconds and then are masked by the Martha's Vineyard ferry getting underway.

Multiple permutations of spectrogram parameters were evaluated in an attempt to optimize the detectibility of divers. The parameters presented in table 3.2 qualitatively optimized a human evaluator's ability to detect the presence of divers. These parameters were kept constant throughout the thesis to maintain uniformity across the data. No filters were applied to the data, allowing evaluation of the entire 0-30 kHz spectrum.

3.4 Manual Evaluation of Diver Data

A data sheet outlining key parameters was recorded for each dive including the times that the divers were at each distance from the hydrophone as well as the time that they entered

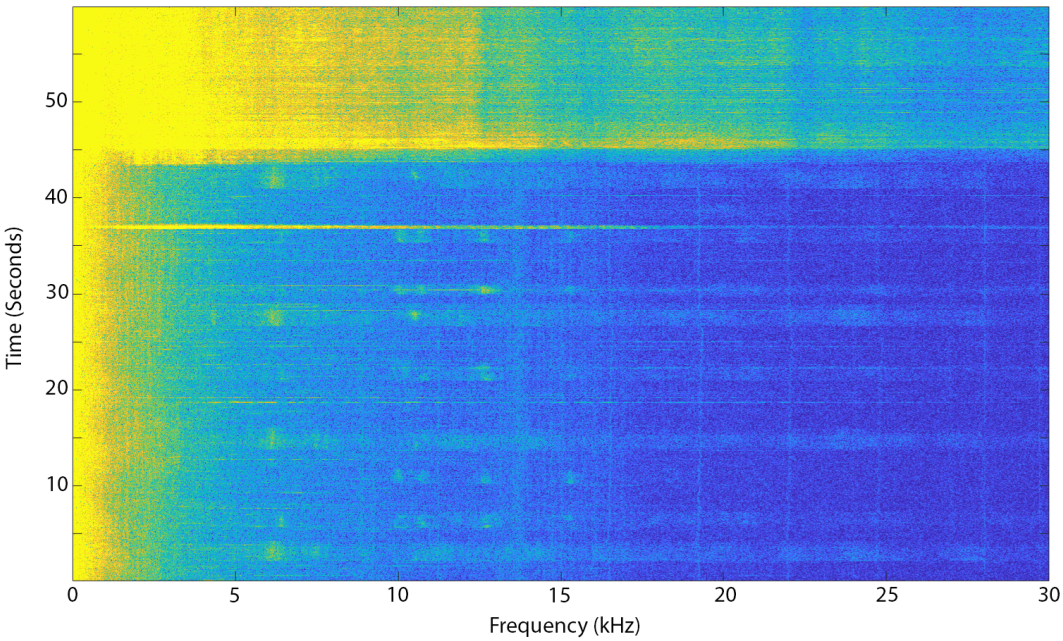


Figure 3-3: Spectrogram from 1013 on 19 October 2018.

Table 3.2: Spectrogram Parameters.

Color Map	Default
Maximum dB	-90
Minimum dB	-140
NFFT	12800 Points
Window	Hamming 6400 Points
Overlap	3200 Points (50%)

and exited the water. Visual evaluation of data enabled reliable identification of the time when the divers arrived at less than 1.5 m from the hydrophone. Times on the data table were adjusted to match the time-base of the acoustic recording. The times provided by the divers and the times identified by manual evaluation often differed slightly, however they were never off by more than three minutes. This adjustment allowed the author to calibrate the data table to the acoustic recordings of the divers, enabling the author to determine the exact range from the divers to the hydrophone during the recording.

Spectrograms were evaluated for the duration of each dive to determine if a trained operator, the author, could detect the divers. If divers were identified the quality of the

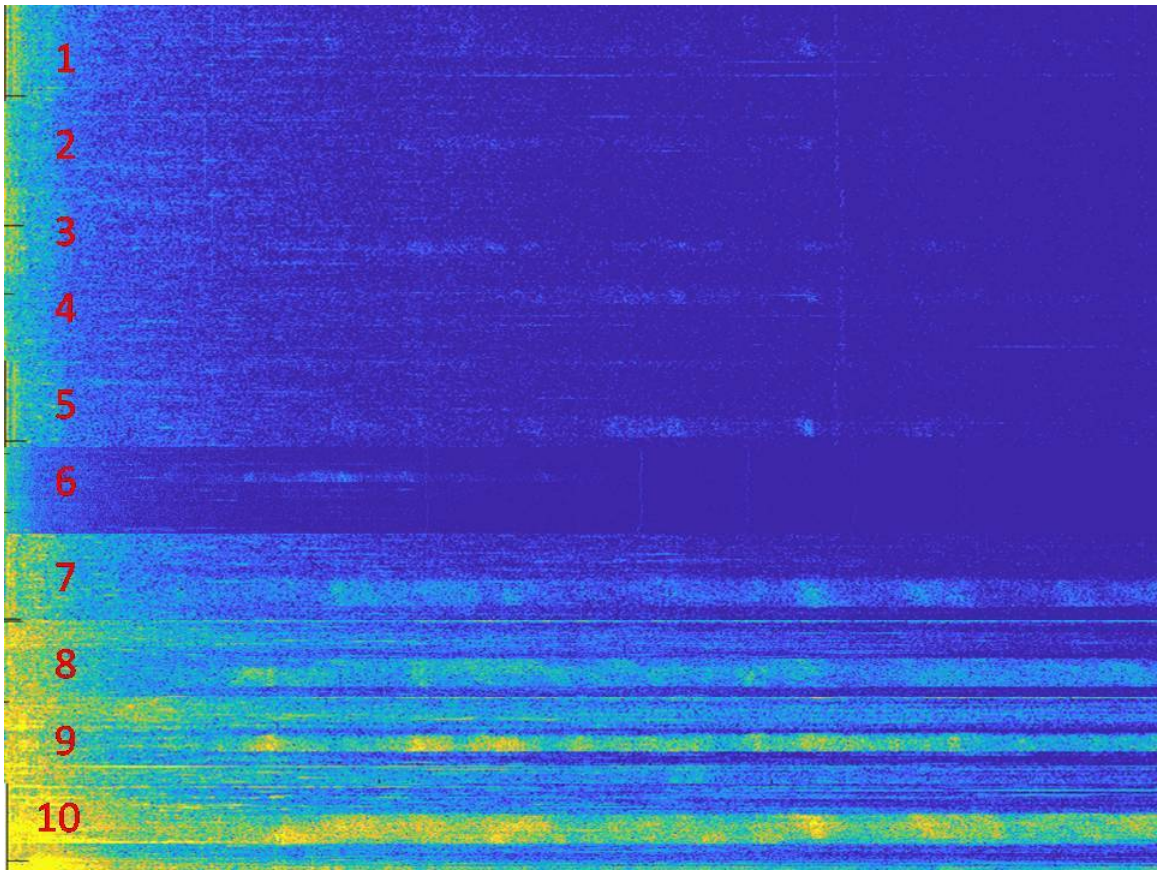


Figure 3-4: Combination of Multiple Spectrograms for Data Labeling. This Figure Was use to Assign a Quantitative Value to Diver Signal Strength During the Manual Data Review Process.

detection was quantified and documented. Appendix C is an example of the form used to record these data. The only human reviewer of the data was the author. It is likely that labeling errors were made; however, the risk of errors was minimized because the author knew when divers were near the hydrophone and when they were not and the author is highly trained and experienced in spectrogram interpretation. If errors were made they would have negatively impacted the network's performance. Figure 3-4 is a combination of multiple spectrograms containing diver emissions with a numerical value correlating to the quality of the emission. This was used as the guide for quantifying the quality of the diver signature.

Analysis showed the quality of diver detection was dependent on the background noise during the dive and the range from the divers to the hydrophone. Figures 3-5 - 3-7 show the dependence of diver detection on both background noise, and distance. Figure 3-5 shows

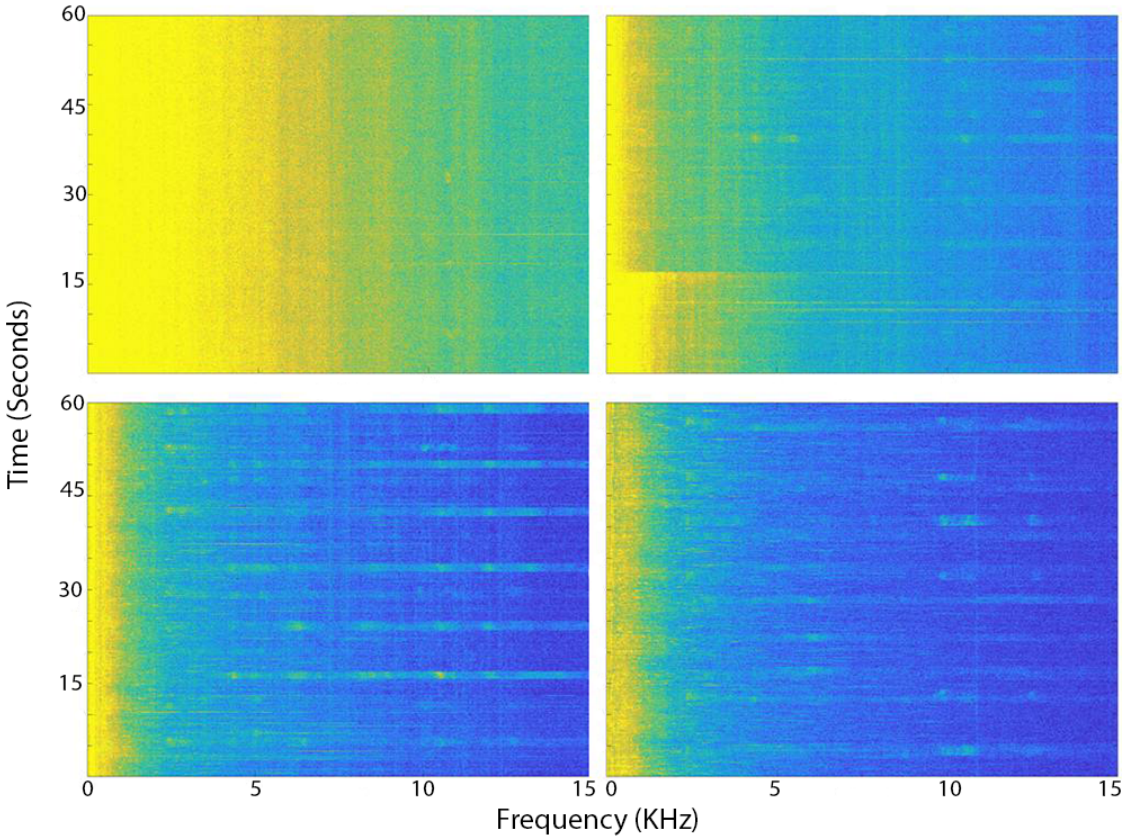


Figure 3-5: Diver Signature at 9.14 m (30 Feet) on Four Days. Top Left: 05 October 2018, Top Right: 19 October 201, Bottom Left: 26 October 2018, Bottom Right: 31 October 2018.

divers at the same distance, 9.14 m (30 feet), across four different days at different background noise levels. Figure 3-6 depicts divers on the same dive, and at different distances, with low background noise. Figure 3-7 is the same as figure 3-6 except it shows a dive with higher background noise.

3.5 Data Labeling and Processing for Machine Learning

Data were prepared for machine learning by band pass filtering it and displaying it in 10 second spectrograms. Initially a single frequency band was chosen, however a second frequency band was later added for better detection in high background noise environments. The data were labeled and manually evaluated for later use in a machine learning network.

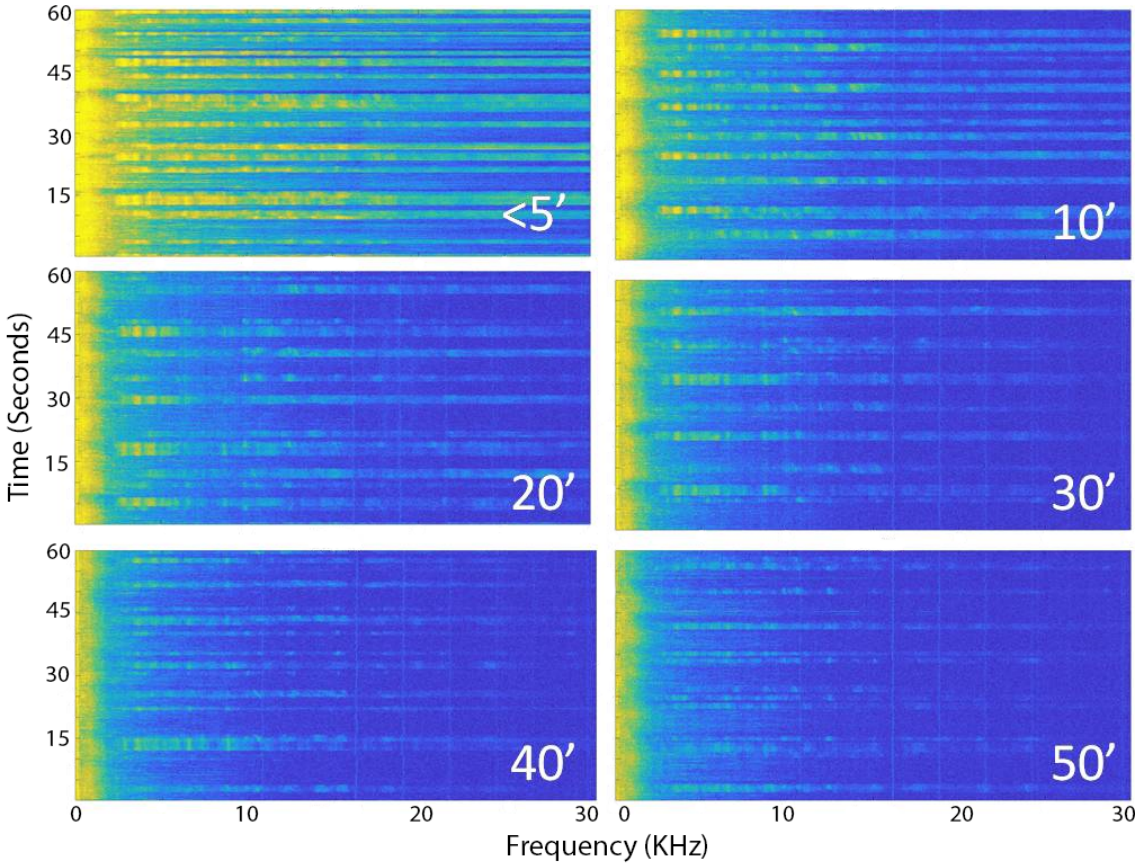


Figure 3-6: Diver Signature as a Function of Range with Low Background Noise. 31 October 2018.

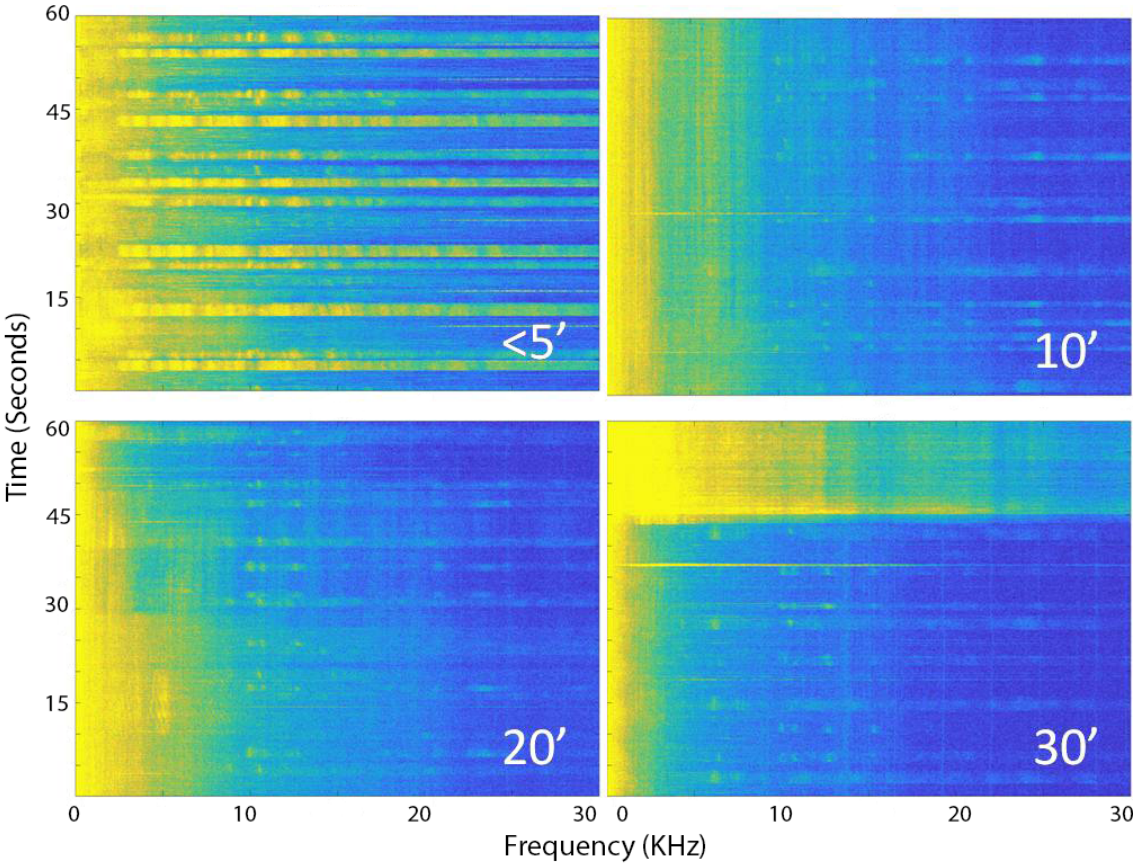


Figure 3-7: Diver Signature as a Function of Range with Moderate Background Noise. 19 October 2018.

3.5.1 Initial Data Labeling and Processing for Machine Learning

An image processing approach for machine learning was chosen because it is a proven method for sonar classification. Additionally, to the knowledge of the author, an image processing, machine learning method for diver detection had not previously been evaluated. Spectrograms were selected as the input for the machine learning network because they display the frequency content of the signal with respect to time. Spectrograms are the same tool used by the Navy for target classification with passive sonar.

Data from the first several dives indicated that the diver signature was most distinguishable in the 10-13 kHz band. For this reason a band pass filter of 7.5-15 kHz with an order of 20 was chosen to filter the acoustic data. The data were displayed from 8-15 kHz. Though this range was generally the best, there were times when humans could detect divers outside of this range as shown in figure 3-8.

Spectrograms with a gray-scale color map and 10 second duration were utilized. The gray-scale color map was used because signal strength is directly proportional to intensity and therefore it was likely well suited for machine learning. The 10 second duration was chosen because it limited the number of diver transients in each spectrogram while maintaining a high likelihood that each would contain a diver emission. This minimized the complexity of individual spectrograms while ensuring they contained sufficient data for diver detection. One full spectrogram was required for diver detection, therefore reducing the spectrogram duration limited the required integration time. It is notable that Zhao *et. al.* also use a 10 second duration for their support vector machine diver detection system [51]. The fact that the sample length selected for this thesis matched the duration chosen by Zhao *et. al.* appears coincidental.

A MATLAB script was used to generate the spectrogram parameters discussed above. The script also cropped the images, to remove labels and titles. It automatically saved the spectrograms in a PNG format, placing applicable information in the file name. The information in the title included the date and time of the data, if there was a diver in the water at the time, the range of the diver, if there was a ship moored at the WHOI pier, the name of the ship, and a serial number representing the parameters used to construct the spectrogram. Appendix D contains a sample spectrogram title along with a key identifying the parameters in the title. Figure 3-9 shows the six, ten second spectrograms in the machine

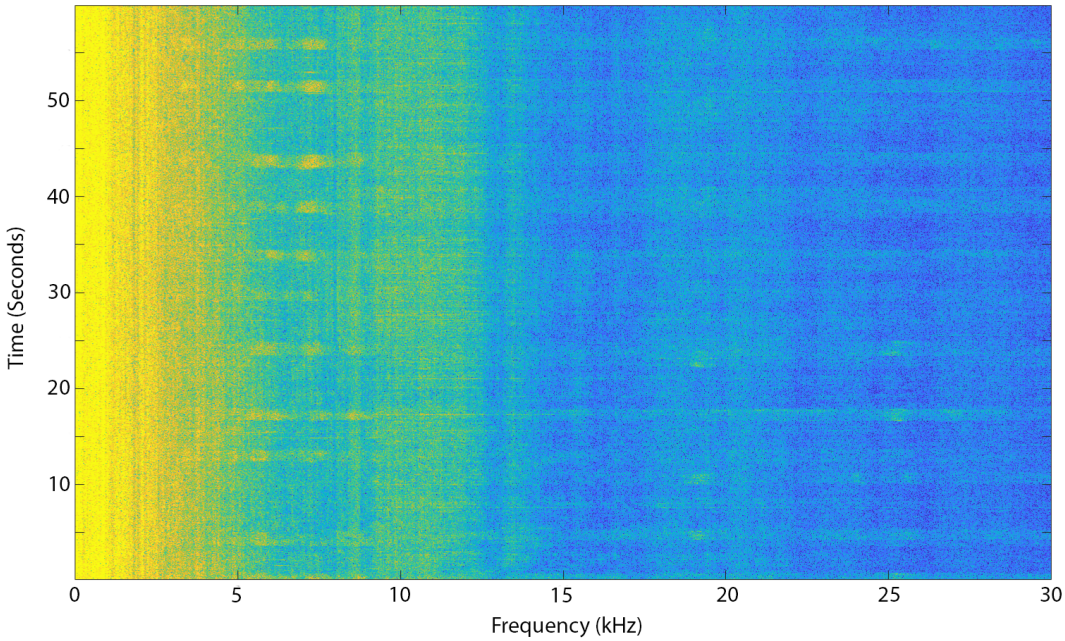


Figure 3-8: Diver Signature 1001 18 December 2018 at a range of 3.04 m (10 Feet). The Bulk of the Diver Signature is Present Outside of the 8-15 kHz Band.

learning format from the 1013 minute of 19 October. Note this is the same data depicted in figure 3-3.

The spectrograms for machine learning were manually evaluated by the author to ensure each one contained a diver transient identifiable by a trained human. The spectrograms with divers in the water but not detectable by a human were segregated for later evaluation. In figure 3-9 the sixth spectrogram, from 50-60 seconds, was segregated because the diver transients were masked by the Martha's Vineyard ferry. The other five spectrograms contained visible diver transients and were used in machine learning.

A control data set was generated by recording just before or after a dive when divers were not in the water. This data set was equal in size to the set containing divers. The data for the control set was within one hour of the dive to replicate the acoustic conditions of the dive as closely as possible. These data complemented the data containing diver transients, providing two data categories, one with divers, and one without.

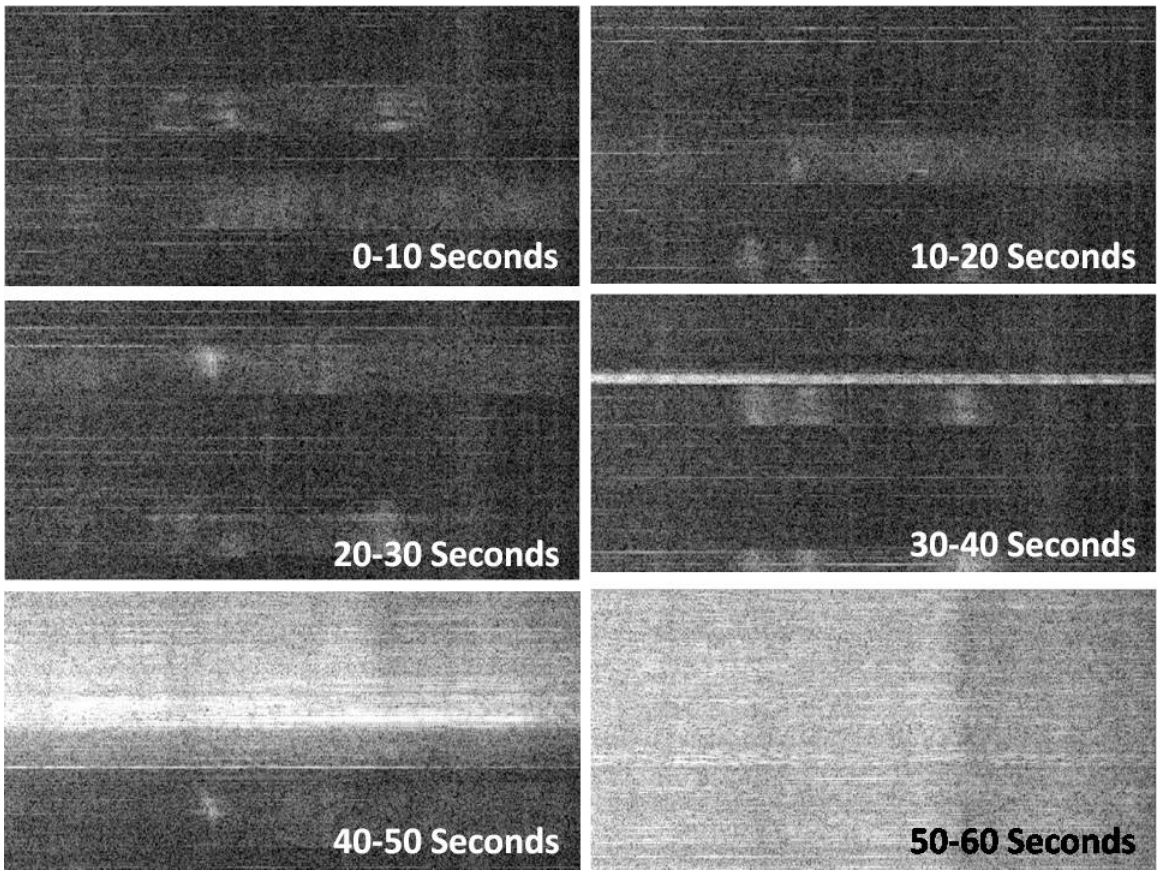


Figure 3-9: Examples of Spectrograms in Machine Learning Format. Diver Range of 9.14 m (30 Feet) on 19 October 2018.

3.5.2 Additional Frequency Band for High Background Noise Environments

Dives conducted on 19 December 2018 and 04 January 2019 contained minimal data where divers were detectable by a human in the frequency band of interest; however, divers were identifiable in higher frequencies. High background noise masked divers in the 8-15 kHz range and therefore a second frequency band of 18-25 kHz was instituted. The low frequency limit of 18 kHz was chosen because there was often a strong diver signature at approximately 19 kHz, and the high frequency limit of 25 kHz was selected to keep the spectrogram dimensions the same as the lower frequency band, making both bands compatible in the machine learning network. Previous data recorded at sufficiently high frequency was reprocessed at the higher frequency band, producing more data for machine learning.

Empirically it was noted that the lower frequency band was better for detection in lower ambient noise conditions, producing longer detection ranges due to less acoustic attenuation. The higher frequency band was better in elevated background noise, as 18-25 kHz was above the majority of the ambient noise near the WHOI dock. This is discussed further in section 5.1.2. Figure 3-10 is a spectrogram from 04 January 2019 where divers were not detectable in the original frequency band but were clearly visible in the higher frequency band.

3.6 Machine Learning

An overview of the machine learning process is shown in figure 3-11. The process began with splitting the data into three groups, training, testing, and validation. A machine learning model was trained using the training data and then tested with the testing data. The model was tuned by adjusting one or more hyperparameters and retraining the network. Hyperparameters are network parameters that are not learned during the training process and therefore must be sent manually prior to training. The training and testing cycle was repeated until an optimal model was produced. The validation data were then used to ensure that the model performed well on data that it had not previously been exposed to. When new data were produced it was first used to validate the model, then split into the initial three groups and the new, larger data set was used to produce a new model.

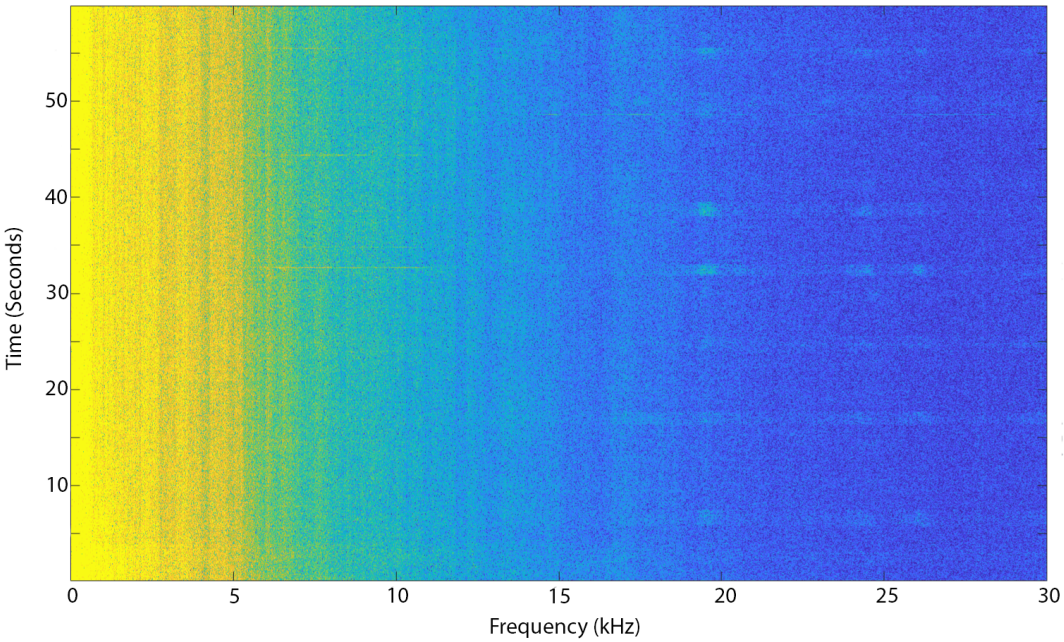


Figure 3-10: Spectrogram with Divers Only Detectable above 15 kHz. 04 January 2019.

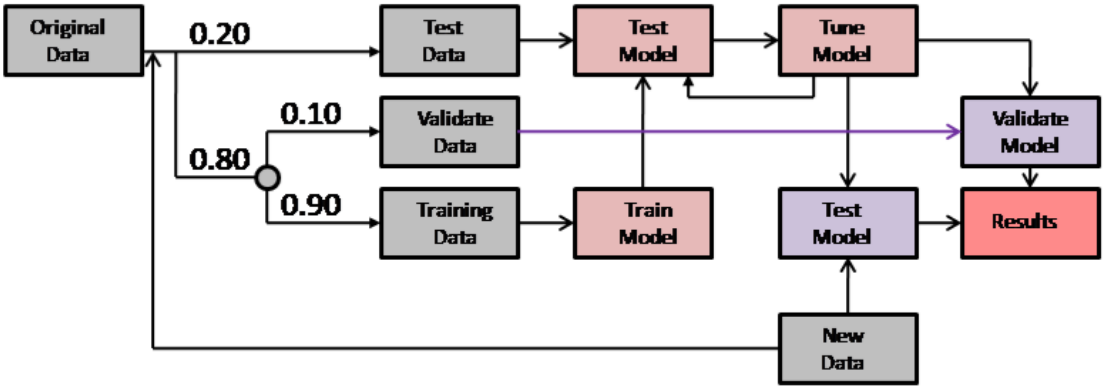


Figure 3-11: Machine Learning Block Diagram. The Data Divide Splits the Data with 20% Becoming Testing Data. The Remaining 80% Becomes Training Data with a 90% Probability and Validation Data with a 10% Probability.

3.6.1 Machine Learning Data Split

Data for machine learning was split into three groups prior to training a model. The groups consisted of training data, testing data, and validation data. The training data were used to train the machine learning model. The testing group was used to evaluate the model's performance during tuning. Following tuning the validation data were employed to verify the model performed well on data not used in the training or tuning process.

The original data set consisted of data from the eleven dives between 5 October 2018 and 30 November 2018. This data set contained 858 10 second spectrograms, half of which contained acoustic emissions from divers and half which did not. All of the data in this set was from the lower, 8-15 kHz, frequency band as the implementation of the higher frequency band occurred later. The intention was to use this data set to generate a preliminary model and to refine this model later when more data were available.

The data were spit such that the bulk was used for training the model. In general, it is recommended that the training set contain between 1,000-5,000 images of each class [36]. This was not possible with the original data set so the bulk of the data were placed in the training group to meet this recommendation as closely as possible. The data split was conducted randomly, using a Python script. Images had an 80% chance of becoming training data and 20% chance of becoming testing data. The validation data were taken from the training data, with a 10% chance that a training spectrogram would be shifted to the validation set and removed from the training set before any training took place. The end result was approximately 72% became training data, 20% testing data, and 8% validation data.

The training data were used to train new models and only used for this purpose. The testing data were used for model tuning which is described in detail in section 3.6.5. During tuning, hyperparameters in the network were adjusted and the model was evaluated against the testing data until optimal performance was achieved. The validation data were used after the model was tuned to ensure that it performed well on data that was not used in the training or tuning process.

As shown in figure 3-11, new data were initially tested using a previous model. This allowed the new data set to be used directly as validation data, evaluating how well the model properly classified new data. After validation the new data were then split with the

same probabilities as the original data set and added to the original data. The original data remained in its originally assigned category of training, testing, or validation. Specifically, once data were classified as training data, it remained training data for this entire process. The same was true for testing and validation data. The new, larger data set was then used to train a new model and the process repeated.

3.6.2 Machine Learning Metric Selection

Several choices for model tuning are available including to maximize the probability of detection, minimize the probability of false alarm, or a combination of thereof. The three primary metrics considered for optimization were precision, recall, and F1 score [39].

Precision is a measure of model's ability to avoid false positives. It is the ratio between number of instances the model correctly identified as positive and the total number of instances the model identified as positive. Alternatively stated, for all instances classified as positive, precision is the fraction actually positive.

$$Precision = \frac{True\ Positives}{Total\ Positives}$$

Recall measures the model's ability to identify all positive instances. It is the ratio of true positives to the sum of true positives and false negatives. Alternatively stated; recall is the fraction of all positive instances classified correctly.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

The F1 score is a weighted harmonic mean between recall and precision where the best score is 1.0 and the worse score is 0.0. F1 is a balance between minimizing false positives and identifying all true positives. F1 score was the metric selected for optimization in the model tuning process.

$$F1 = 2\ \frac{Precision * Recall}{Precision + Recall}$$

The choice to maximize F1 score was based on the lack of a specific penalty structure for this work. If there was an objective function, the choice of optimization may have been different. If the objective was detecting divers every time possible, independent of false alarm rate, recall would have been selected. Conversely, if there was a high penalty for false alarms compared to missed detection, precision would have been chosen. Taken to the

extreme either of these choices would have resulted in classifying every image as divers or every image as non divers. This would have resulted in no missed detections, or no false alarms respectively, but the model would have provided no value. With no clear objective function to optimize, maximizing F1 score appeared the most appropriate choice [41].

3.6.3 Machine Learning Software and Packages

Machine learning in this thesis was conducted in Python. Python was chosen because it is the programming language with the largest assortment of libraries designed specifically for machine learning [36]. The following python packages were used in the machine learning network presented later in the thesis. Figure 3-12 shows the inter-dependencies of the Python packages used in this thesis.

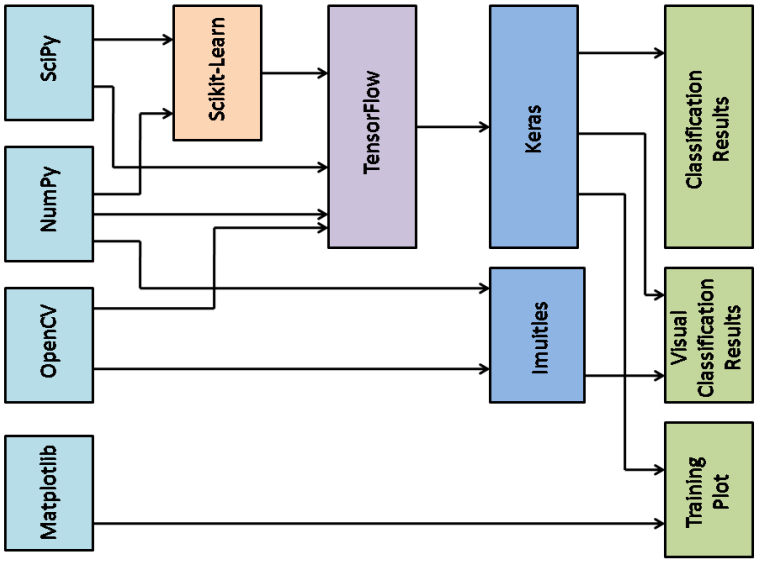


Figure 3-12: Python Package Inter-Dependencies.

Google TensorFlow

Google TensorFlow is a computational framework for building machine learning models [4]. It makes machine learning simpler and faster to implement [4]. TensorFlow is written with a python front end, simplifying programming, but conducts execution in C++ for better performance. The key benefit of TensorFlow is abstraction, allowing the programmer to focus on the higher-level machine learning implementation as opposed to the lower level

programming required for machine learning [50][5].

Keras

Keras is an application program interface (API) for high-level neural networks. Like TensorFlow it is written in Python. A separate machine learning framework such as CNTK, Theano, or TensorFlow is required for Keras to function. TensorFlow was used as the back-end for Keras in this thesis. Keras was designed to be modular and allow for fast prototyping of neural networks. It supports both convolutional and recurrent neural networks, and can be run on either a central processing unit or a graphics processing unit [1][48].

Scikit-Learn

Scikit-learn is a machine learning library for classification, regression and clustering algorithms. In this thesis it was used for classification, determining the presence or absence of divers. Scikit-learn is written in Python. It requires several other packages including NumPy, SciPy, and Matplotlib [32].

Open Source Computer Vision Library (OpenCV)

OpenCV is a computer vision and machine learning software library that was used in this thesis for image processing tasks in machine learning. It was designed for real time computer vision. OpenCV has become *de-facto* standard for image processing in machine learning in recent years [36].

Imutils

Imutils is an image processing library that works with OpenCV. It was written by Dr. Adrian Rosebrock from PyImageSearch and provides functions for image processing [35]. In this thesis it was used for data augmentation, artificially increasing the size of the training data set, and to display results of model evaluation. Data augmentation is discussed in section 3.6.5.

Matplotlib

Matplotlib is a plotting library for python that produces two dimensional plots [2]. Matplotlib was used to plot the training process which for evaluation of over-fitting.

SciPy

Scipy is a set of numerical and scientific tools for Python. It is written in python but outputs to C++ binaries for more efficient execution. SciPy is used by several of the packages listed above [3].

Numpy

Numpy is a python library for multi-dimensional arrays. It enables Python to conduct efficient operations on arrays of data. Numpy is used by several of the packages listed above [3].

3.6.4 Network Overview

A deep convolutional neural network was used in this thesis. A straight-forward network was chosen for use on a standard laptop with minimal training time. Figure 3-13 is a graphical depiction of the network. The network was built using a Keras architecture. This convolutional neural network was similar in structure to a network used by Dr Adrian Rosebrock from PyImageSearch to evaluate root health of hydroponic plants [37]. That network was loosely based on AlexNet and OverFeat [41][25].

The network used in this thesis consisted of three modules of two dimensional convolutional layers followed by ReLu activation layers, pooling layers, and dropout layers. A module containing a fully connected layer, a ReLu activation layer, and dropout followed. The network was concluded with a second fully connected layer and a softmax classification layer. This architecture was chosen as it was straight-forward enough to be run on a commercial laptop while leveraging a modern deep convolutional neural network framework and using the most common activation function [27].

The convolutional layers extracted features from their inputs [34]. The initial convolutional layer extracted low level features from the image, but subsequent layers identified increasingly complex features [27]. The input of the first convolutional layer was a raw image. Subsequent convolutional layers used the activation map produces by the previous module as their input. Each convolutional layer had a receptive field which scanned the input. The scan was convolved with learned weights to extract features and create a feature map [27]. The feature map was smaller in dimension than the input as a result of the finite

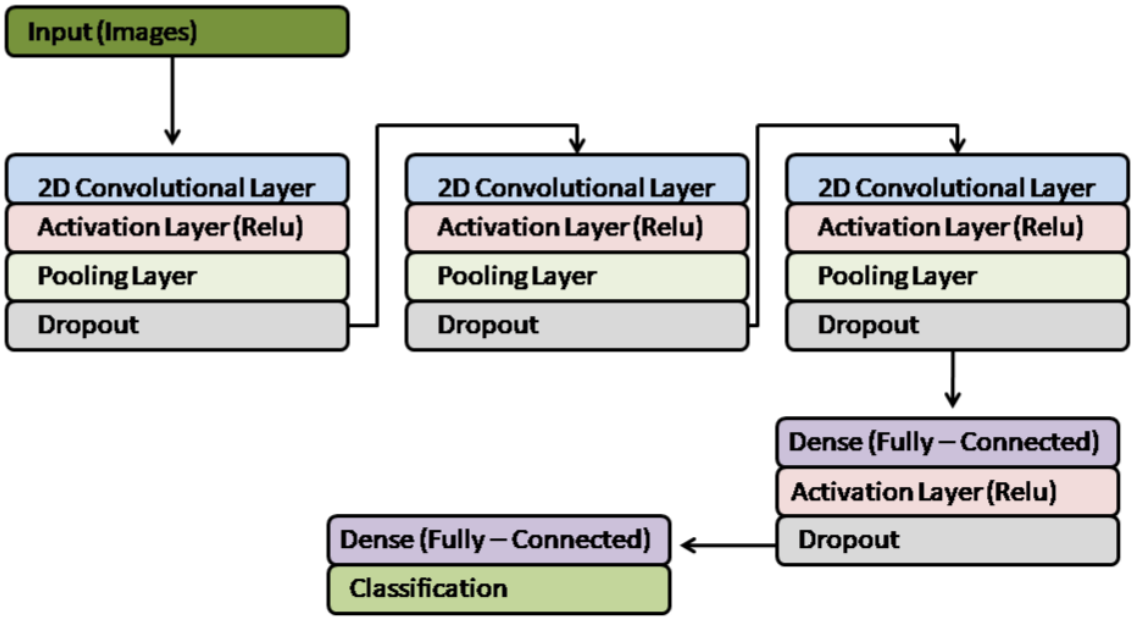


Figure 3-13: Block Diagram of the Deep Convolutional Neural Network Used.

size of the receptive field used to scan the input. The convolutional neural network used in this thesis had a receptive field size of 7x7 for the first convolutional layer, 5x5 for the second, and 3x3 for the third.

A nonlinear activation ReLu was applied following each of the convolutional layers and the first fully connected layer. This layer increased the nonlinear properties of the previous layer and sets all negative values to zero. This allowed the network to extract non-linear features from the feature map [34]. The output of the activation layer was an activation map.

A pooling layer followed each of the activation layers. Pooling layers were effectively down-sampling layers, which took an input of neuron clusters of a given dimension from the activation map and output a single number. This resulted in spatial invariance to translations or distortions of the input [34]. In this thesis max pooling layers were used with an input size of 2x2. Max pooling used the highest of the input numbers as the output.

A dropout layer was included at the end of each module. The dropout layer set a random group of activations to zero. This is a form of regularization and helped minimize over-fitting [19]. Two sets of fully connected layers were included towards the end of the neural network. These layers connected every neuron in the previous layer to each neuron in the next layer.

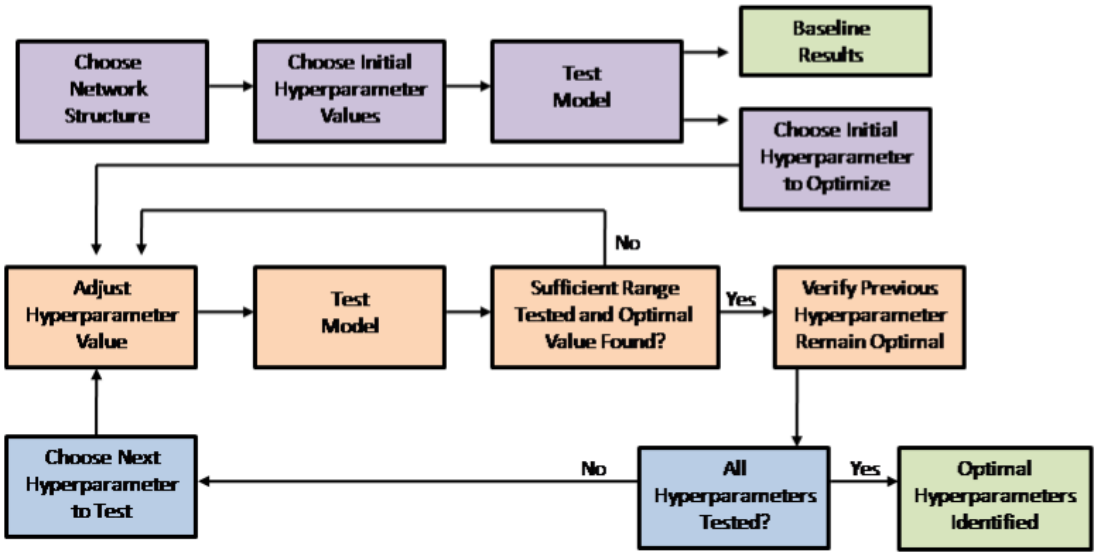


Figure 3-14: Model Tuning Flow Chart.

The fully connected layers served as a feature extractor for higher level reasoning [34].

The last layer in the neural network was a softmax classification layer. The classification layer produced a probability distribution for the possible categories; diver or non-diver. This probability distribution was used to produce the final classification of each image.

3.6.5 Model Tuning

Model tuning followed network development. A model was trained with the training data set, then evaluated with the testing data set. One or more hyperparameters were then adjusted and the training and testing process repeated. This cyclic process was continued until a model with optimal performance was found. Two discrete models were developed during model tuning. The first, single frequency model, was produced from the 11 dives trough 30 November 2018, using data from the 8-15 kHz band only. The second, dual frequency model, was constructed from the 15 dives through 9 January 2019, using data from both the 8-15 kHz and 18-25 kHz frequency bands. Current best practice for all machine learning requires considerable tuning. This is one of the limitations of machine learning. Figure 3-14 is a graphical depiction of the model tuning work flow. Table 3.3 lists the key hyperparameters tuned for the network used in this thesis.

Table 3.3: Hyperparameters Adjusted During Model Tuning.

Hyperparameter	Description
Learning Rate	Step size the model takes in the direction of the apparent minimum during training.
Learning Rate Decay	Amount the learning rate is lowered following each epoch during training.
Data Augmentation	A form of regularization: Manipulation of the original training data set to create an artificially large training set.
Regularization Constant	Penalty for large weights in the learning process. Minimizes overfitting and helps the model generalize.
Dropout	A form of regularization: Randomly disconnecting a portion of the connections from the preceding layer.
Number of Epochs	Number of times the training algorithm is exposed to each piece of training data.
Activation function	Type of non-linearity used.

Tuning Individual Hyperparameters

Learning rate tends to be the most dominate hyperparameter and therefore it is often tuned first [36]. The optimal learning normally rate falls between $1 * 10^{-5}$ and $1 * 10^{-3}$ [29]. If the learning rate is too large the network will alternate around a local or global minimum, never finding the minimum. If the learning rate is too small, the step size will be too small, and again, a global or local minimum may not be reached. The initial learning rate for the single frequency model was set at $1 * 10^{-4}$ to evaluate the models performance with an intermediate value. This learning rate resulted in an average F1 score, between divers and no divers, of 0.85. Testing a learning rate of $1 * 10^{-3}$ produced in an average F1 score of 0.83 and testing $1 * 10^{-5}$ generated an average F1 score of 0.72. Multiple other learning rates were tested indicating the optimal learning rate was between $8 * 10^{-5}$ and $3 * 10^{-4}$. Further testing suggested that the optimal learning rate with other hyperparameters held constant was $3 * 10^{-4}$. Table 3.4 and figure 3-15 show the results of all learning rates evaluated for the single frequency model.

Figure 3-16 shows Training Loss and Accuracy on the single frequency model for a learning rate of $3 * 10^{-4}$. Note there is not a large gap between training accuracy and validation accuracy, indicating that over fitting is not occurring in the model.

Table 3.4: Tuning Learning Rate: Single Frequency Model.

Learning Rate	Weighted Average F1 Score	Divers F1 Score	No Divers F1 Score	Order tested
$1*10^{-5}$	0.72	0.64	0.79	3
$5*10^{-5}$	0.81	0.78	0.83	6
$8*10^{-5}$	0.85	0.84	0.86	7
$9*10^{-5}$	0.83	0.82	0.83	11
$1*10^{-4}$	0.85	0.83	0.86	1
$2*10^{-4}$	0.85	0.84	0.86	10
$3*10^{-4}$	0.85	0.85	0.86	9
$5*10^{-4}$	0.85	0.84	0.85	5
$7*10^{-4}$	0.29	0.63	0	8
$1*10^{-3}$	0.83	0.81	0.85	2
$1*10^{-2}$	0.29	0.63	0	4

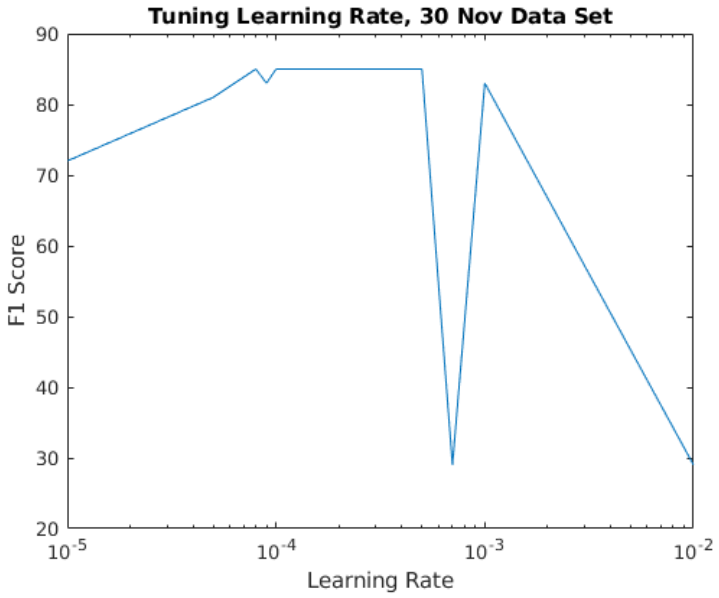


Figure 3-15: Tuning Learning Rate: Single Frequency Model. Model Performance is Best with a Learning Rate Between $8*10^{-5}$ and $5*10^{-4}$. There is a Local Maximum of the Loss Function at $7*10^{-4}$ and a Local Minimum at $1*10^{-3}$.

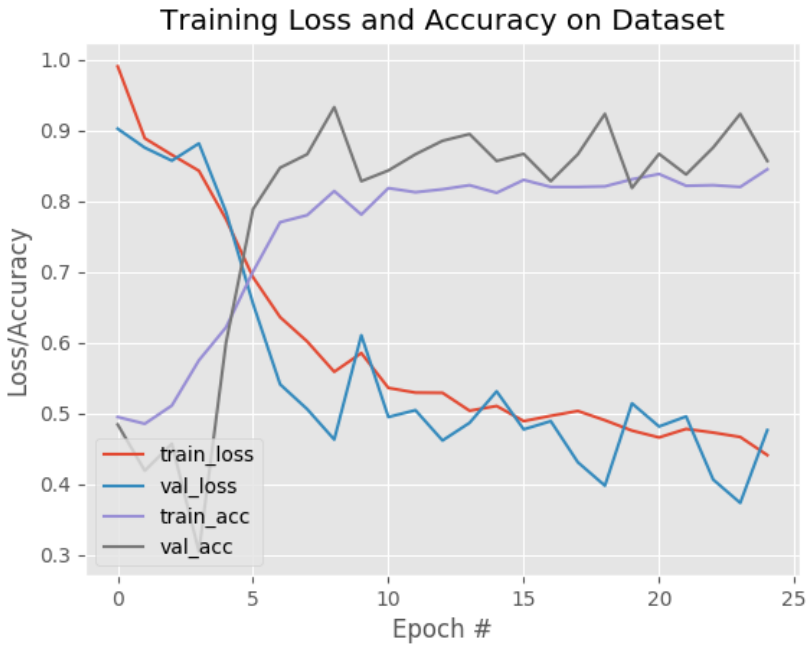


Figure 3-16: Training Loss and Accuracy During Training: Single Frequency Model.

With learning rate established at $3 * 10^{-4}$, learning rate decay was adjusted to determine if the original decay structure was optimal. Learning rate decay reduced the learning rate each epoch, forcing the network to take smaller steps as the loss function minimum was approached. Initially learning rate decay was set as learning rate divided by the number of epochs tested. For a learning rate of $3 * 10^{-4}$ and 25 epochs, after each epoch the learning rate was reduced by $12 * 10^{-6}$. Several other decay structures, including no decay, were tested but underperformed the original decay structure.

$$Learning\ Rate\ Decay\ = \frac{Learning\ Rate}{Number\ of\ Epochs}$$

Regularization is generally considered to be second only to learning rate in importance and was tuned next [36]. Regularization consists of several hyperparameters including the regularization constant, dropout, and data augmentation. Data augmentation was the first regularization hyperparameter evaluated. Several alternative augmentation schemes, including no augmentation, were evaluated and it was determined that the baseline data augmentation scheme performed best. The baseline data augmentation scheme is shown in table 3.5. Of note implementing a vertical image flip along with a horizontal flip was

Table 3.5: Baseline Data Augmentation Scheme.

Rotation Range	± 20%
Zoom Range	5%
Width Shift Range	5%
Height Shift Range	5%
Horizontal Flip	TRUE
Vertical Flip	FALSE

detrimental so only the horizontal flip was used. It was likely the features of a scuba diver had a vertical dependence, with frequency content changing with time, and therefore the vertical flip made it harder for the network to distinguish the proper features of a scuba diver. The first reported use of data augmentation was by LeCun *et. al.* in 1998 and is now best practice [26][36].

The regularization constant was tuned next. Several iterations revealed that the optimum value for the regularization constant was $2 * 10^{-4}$. Dropout, the final hyperparameter for regularization was then tuned. Tuning indicated the optimal setting for dropout was 0.10 for each dropout layer. Higher dropout lead to lower performance due to too high a percentage of the network being disconnected. Lower levels of dropout resulted in over fitting and ultimately lower performance. Table 3.6 shows the results of all permutations of regularization considered. Figure 3-17 depicts regularization constant tuning with constant dropout.

During model tuning for learning rate, regularization constant, and dropout the model performance degraded significantly at both extremes of the values tested, with an F1 score at least 0.1 below the optimal. Between the extreme there was a wide range of values that had nearly constant performance. This indicated within a given range of values, roughly an order of magnitude for learning rate and regularization constant, the network was not sensitive to the values of the hyperparameters. The other hyperparameters tested, including the activation function and data augmentation were significantly more sensitive with non optimal choices resulting in an F1 score at least 0.1 below the optimal. These hyperparameters were a series of discrete choices, as opposed to continual values, and were therefore more difficult to evaluate for sensitivity. It is not surprising that a ReLu activation function with data augmentation performed best as these normally produce the best results [27][36].

Table 3.6: Tuning Regularization Constant and Dropout: Single Frequency Model.

Regularization Constant	Dropout Following Pooling	Dropout Following Activation	Weighted Average F1 Score	Divers F1 Score	No Divers F1 Score	Order tested
$5*10^{-5}$	0.25	0.5	0.76	0.73	0.79	10
$1*10^{-4}$	0.25	0.5	0.85	0.83	0.87	9
$2*10^{-4}$	0.25	0.5	0.85	0.86	0.85	1
$4*10^{-4}$	0.25	0.5	0.8	0.79	0.81	2
$6*10^{-4}$	0.25	0.5	0.83	0.81	0.84	3
$8*10^{-4}$	0.25	0.5	0.82	0.8	0.83	4
$1*10^{-3}$	0.25	0.5	0.82	0.81	0.84	5
$1.2*10^{-3}$	0.25	0.5	0.82	0.78	0.85	6
$1.5*10^{-3}$	0.25	0.5	0.82	0.8	0.83	7
$2*10^{-3}$	0.25	0.5	0.55	0.37	0.7	8
$2*10^{-4}$	0.3	0.6	0.61	0.69	0.55	15
$2*10^{-4}$	0.15	0.3	0.84	0.83	0.86	11
$2*10^{-4}$	0.1	0.2	0.85	0.84	0.86	12
$2*10^{-4}$	0.1	0.1	0.86	0.84	0.86	13
$1*10^{-4}$	0.1	0.1	0.81	0.8	0.83	14

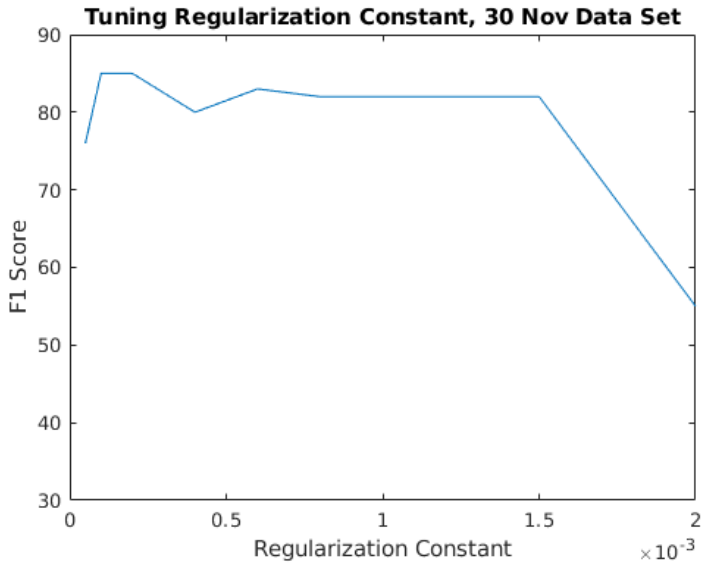


Figure 3-17: Tuning Regularization Constant: Single Frequency Model.

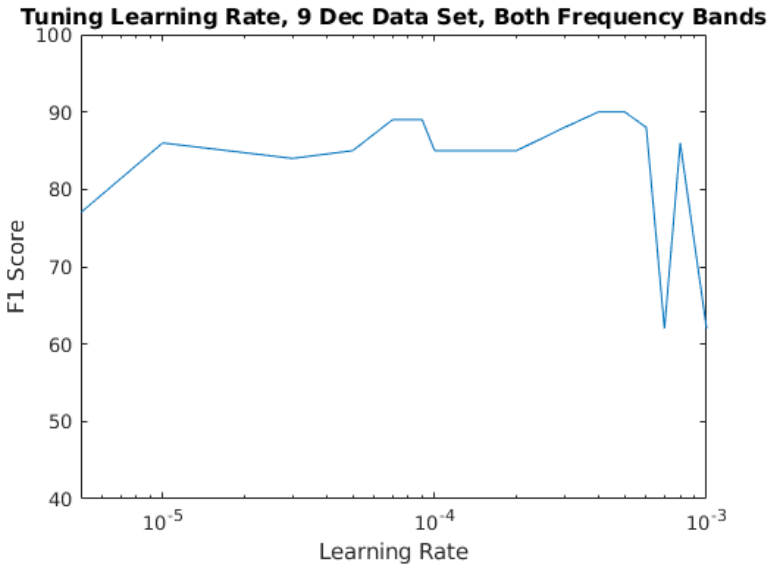


Figure 3-18: Tuning Learning Rate: Dual Frequency Model.

Once the regularization hyperparameters were set the number of epochs in the network was shifted from 25 to 100. This resulted in slightly increased performance, on the order of 0.01 or 0.02 per model but at the cost of a 4x increase in training time. Several variants of learning rate and the regularization hyperparameters were evaluated with 100 epochs. The learning rate and regularization scheme outlined above remained optimal.

Model Validation and Continued Tuning with New Data

Four dives were conducted between 19 December 2018 and 9 January 2019. Because these dives occurred after the single frequency model was generated they were not used to develop or tune it. As a result, they were used to validate the model's performance on new data. The results of this analysis are shown in chapter 4.

The new data were then split using the same statistical breakdown as the original data. This time when the data were split both frequency bands, 8-15 kHz and 18-25 kHz, were used for both the new and previous data. The new training, testing, and validation data were added to the previous data of the same group to maintain the integrity of the original data set. The new training and testing data sets were then used to train and tune a new model which is referred to as the duel frequency model for the remainder of this thesis. The results of tuning are shown in tables 3.7 and 3.8 and figures 3-18 and 3-19.

Table 3.7: Tuning Learning Rate: Dual Frequency Model.

Learning Rate	Weighted Average F1 Score	Divers F1 Score	No Divers F1 Score	Order tested
$5*10^{-6}$	0.77	0.72	0.82	4
$1*10^{-5}$	0.86	0.85	0.86	3
$3*10^{-5}$	0.84	0.85	0.84	5
$5*10^{-5}$	0.85	0.86	0.84	2
$7*10^{-5}$	0.89	0.88	0.89	6
$9*10^{-5}$	0.89	0.88	0.89	7
$1*10^{-4}$	0.85	0.86	0.84	8
$2*10^{-4}$	0.85	0.86	0.85	9
$3*10^{-4}$	0.88	0.88	0.89	11
$4*10^{-4}$	0.9	0.9	0.91	12
$5*10^{-4}$	0.9	0.89	0.9	10
$6*10^{-4}$	0.88	0.87	0.89	13
$7*10^{-4}$	0.62	0.49	0.73	14
$8*10^{-4}$	0.86	0.85	0.86	15
$1*10^{-3}$	0.62	0.48	0.75	1

Table 3.8: Tuning Regularization Constant: Dual Frequency Model.

Regularization Constant	Weighted Average F1 Score	Divers F1 Score	No Divers F1 Score	Order tested
$1*10^{-5}$	0.86	0.84	0.87	3
$5*10^{-5}$	0.89	0.88	0.9	4
$8*10^{-5}$	0.87	0.88	0.87	5
$1*10^{-4}$	0.9	0.89	0.91	2
$1.5*10^{-4}$	0.89	0.87	0.9	6
$2*10^{-4}$	0.9	0.9	0.91	1
$2.5*10^{-4}$	0.9	0.89	0.9	7
$3.5*10^{-4}$	0.88	0.88	0.89	12
$5*10^{-4}$	0.86	0.86	0.85	8
$8*10^{-4}$	0.88	0.88	0.87	9
$1*10^{-3}$	0.86	0.84	0.87	10
$2*10^{-3}$	0.88	0.87	0.89	11

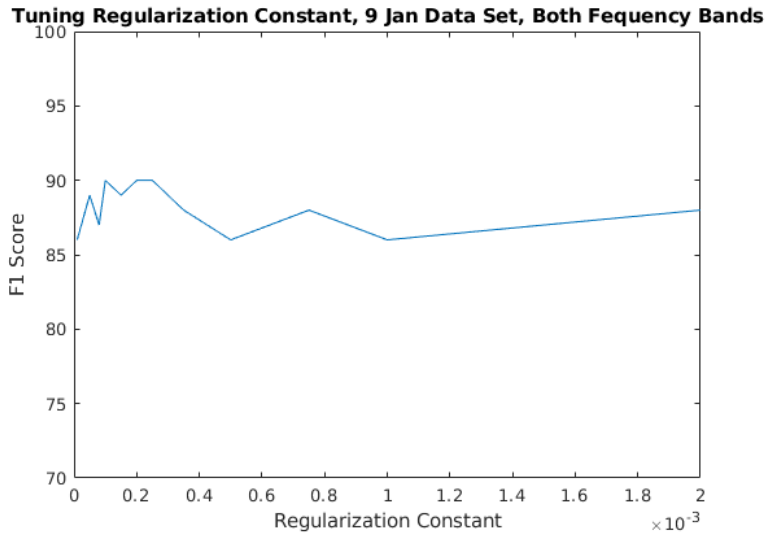


Figure 3-19: Tuning Regularization Constant: Dual Frequency Model.

Table 3.9: Model Performance on Testing Data: Dual Frequency Model.

Weighted Average F1 Score	Divers F1 Score	No Divers F1 Score
0.92	0.92	0.93

Several other parameters were adjusted including the number of epochs and the type of activation function. Table 3.9 shows the results of the dual frequency model evaluated with the testing data. Eighty four different models were trained during the tuning process to produce the final model. This model appeared to be at the minimum of the loss function, resulting in the best results against the training data. Table 3.10 shows the key hyperparameters for this model.

Table 3.10: Final Dual Frequency Model Hyperparameters.

Learning Rate	$4*10^{-4}$
Regularization Constant	2.00E-04
Dropout	0.1 (all layers)
Epochs	100
Activation Function	Relu
Training Time	About 4 Hours

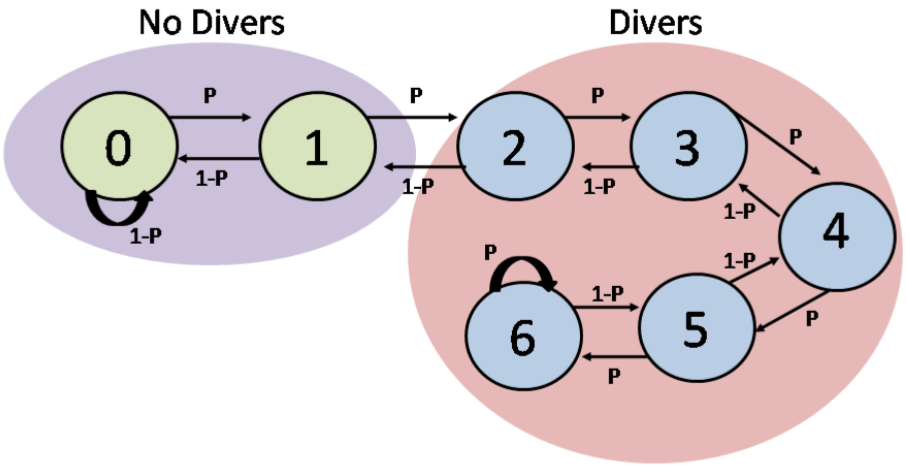


Figure 3-20: Markov Chain Model for Sequential Data Processing. The Classes of No Divers and Divers are Represented by the Purple and Tan Ovals. The Markov States are Represented by the Green and Blue Circles.

3.6.6 Sequential Data Processing

The use of sequential data processing was evaluated to determine if better results could be achieved by including recent results. The idea was that if there was a diver detection just before and just after a missed detection, the system should not be penalized for the missed detection. Likewise, a single false detection would only penalize the system if the detection threshold was set low enough. Processing data in a sequential fashion is essentially creating a filter to remove the "noise" in automated diver detection. The process of using sequential data processing for diver detection was represented using a Markov chain.

Markov Chain Model for Diver Detection

The model for sequential data processing used in this thesis can be modeled using a discrete time Markov chain [7]. The chain had two classes and seven states. All states in the model were recurrent, meaning that it was always possible to get from any state to any other state given enough time steps. Figure 3-20 is a graphical depiction of the Markov chain used for sequential data processing. In this model a classification of divers was reported when the chain was at state two or higher. The choice of state two as the detection threshold was a balance between minimizing integration time required for detection and filtering false detections. This is discussed in detail in chapter 4.

P was the probability that the machine learning model indicated diver detection at any discrete time step of 10 seconds. P was equal to the probability of diver detection (P_D) plus $1 - P_D$ multiplied by the probability of false alarm (P_{FA}).

$P = P_D + (1 - P_D)P_{FA}$

P_D was a function of the detection system used, including hydrophone and machine learning model, the divers range, diver acoustic characteristics, and ambient noise levels. P_{FA} was a function of the detection system and ambient noise. Both P_D and P_{FA} continually changed as diver range and ambient noise changed. Figure 3-21 is the transition probability matrix for the discrete time Markov chain used for sequential data processing.

P_{00}	P_{01}	P_{02}	P_{03}	P_{04}	P_{05}	P_{06}
P_{10}	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	P_{16}
P_{20}	P_{21}	P_{22}	P_{23}	P_{24}	P_{25}	P_{26}
P_{30}	P_{31}	P_{32}	P_{33}	P_{34}	P_{35}	P_{36}
P_{40}	P_{41}	P_{42}	P_{43}	P_{44}	P_{45}	P_{46}
P_{50}	P_{51}	P_{52}	P_{53}	P_{54}	P_{55}	P_{56}
P_{60}	P_{61}	P_{62}	P_{63}	P_{64}	P_{65}	P_{66}

1-P	P	0	0	0	0	0
1-P	0	P	0	0	0	0
0	1-P	0	P	0	0	0
0	0	1-P	0	P	0	0
0	0	0	1-P	0	P	0
0	0	0	0	1-P	0	P
0	0	0	0	0	1-P	P

Figure 3-21: Markov Chain Transition Probability Matrix.

The only two states with a self transition probability greater than zero were states zero and six. State six remained at state six with probability P, as no higher state existed. State zero remained at state zero with the probability $1-P$, as there was no lower state. All other states either transitioned to the next higher state with probability P, or the next lower state with probability $1 - P$.

Data Processing with Markov Chain

Data from four dives on 30 January, 6 February, 19 February, and 27 February 2019 was used for sequential data processing with a Markov chain. These data were previously not seen by the machine learning model used for analysis. The machine learning model used in sequential processing was the dual frequency model presented in section 3.6.5. Ten second spectrograms for both the 8-15 kHz and the 18-25 kHz were evaluated sequentially by the machine learning model. The model's classification of both frequency band at each 10 second interval was placed in a matrix. The model's results for the two-frequency band of the same time were then combined using an or operator, $\bigcup$. This resulted in a positive classification for any time period where either frequency band identified divers.

In addition to the automated classification, a hand classification was conducted on the same data by the author using the full 0-30 kHz spectrogram. The results were placed in the same matrix as the learning model's output. This process was repeated for a period of time when divers were not in the water. The data when the divers were not in the water was equal to the dive duration and taken just prior to the dive.

A Markov chain was used to determine the confidence of diver detection. The same Markov chain was used for both manual and machine learning detections. The chain began at state zero and for every sequential interval that a diver was detected the chain moved to the next higher state, up to a maximum of state six. For every interval that a diver was not detected the chain moved to the next lower state, to a minimum of state zero. State six was chosen as the maximum because one minute from the last diver detection the system would no longer report the presence of divers. Due to the slow speed of divers it was reasonable to assume that after the last detection the divers remained relatively close to the hydrophone for at least one minute.

The Markov chain for diver detection had two classes, one indicating diver detection and indicating the absence of divers. The states contained in the classes were determined by the desired confidence level for diver detection. Low confidence diver detection, shown in figure 3-22, included all states greater than or equal to state one. Medium confidence diver detection, shown in figure 3-21 above, indicated diver detection at state two and above. High confidence diver detection, shown in figure 3-23, included state three and higher for diver detection. This thesis used medium confidence for diver detection. This is discussed

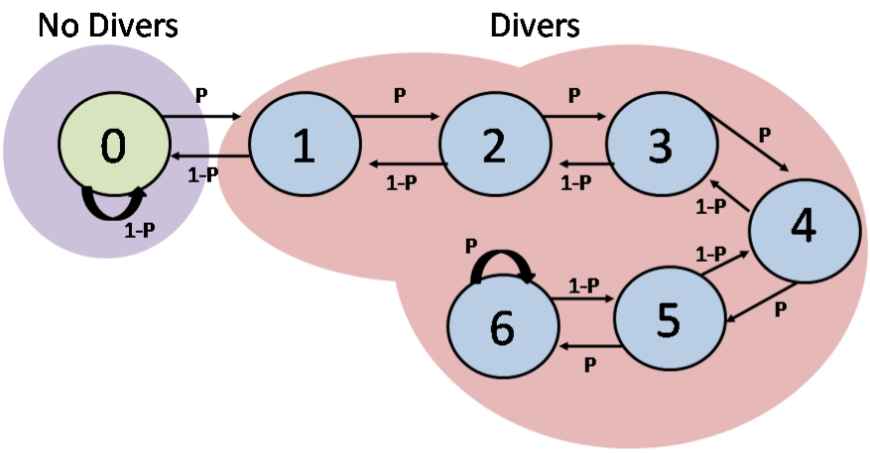


Figure 3-22: Markov Chain Model for Low Confidence Diver Detection.

in detail in chapter 4.

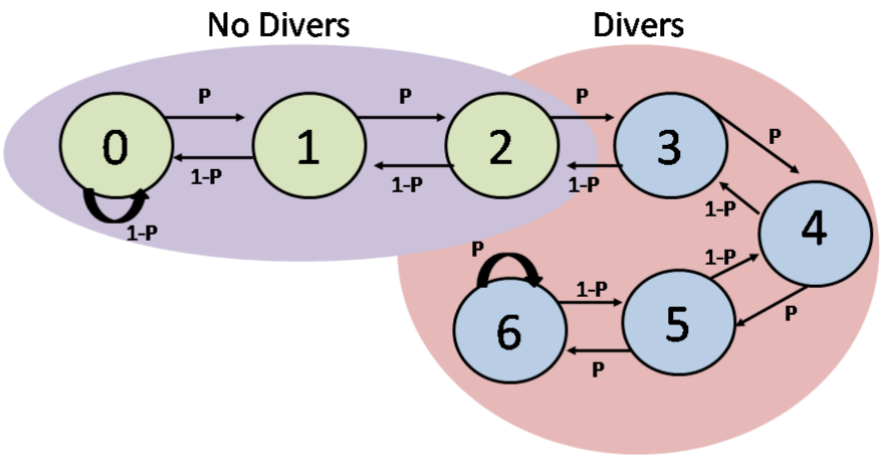


Figure 3-23: Markov Chain Model for High Confidence Diver Detection.

3.7 Alternate Methods of Diver Detection

Other methods for diver detection were also evaluated. An attempt was made to use match filtering by comparing the frequency content of the acoustic signal to the frequency content of a known diver inhale transient. This analysis was conducted with divers at several ranges over several dives. Analysis was also conducted while divers were not in the water but with construction at the Martha's Vineyard ferry terminal occurring. While it would have been possible to use match filtering for diver detection based on the frequency of diver inhalation transients. This method has many shortcomings when compared to machine learning. This type of analysis and its limitations are discussed extensively in section 2.1 and was not replicated in this thesis.

3.7.1 Match Filtering of Diver Inhale Transient

Match filtering of diver inhalation transients was evaluated as a potential method to improve data processing. This was started by identifying a portion of data where the diver signature was very strong with little background noise. Dive 0336 of 31 October 2018 contained data meeting these requirements and contained transients from three distinct divers. One second of unfiltered data was taken from the transient of each diver. This signal was used as the reference signal in match filtering. Figure 3-24 shows spectrograms containing divers at a range less than 1.5m to a range of 9.14 m (30 feet) from the hydrophone. These data were

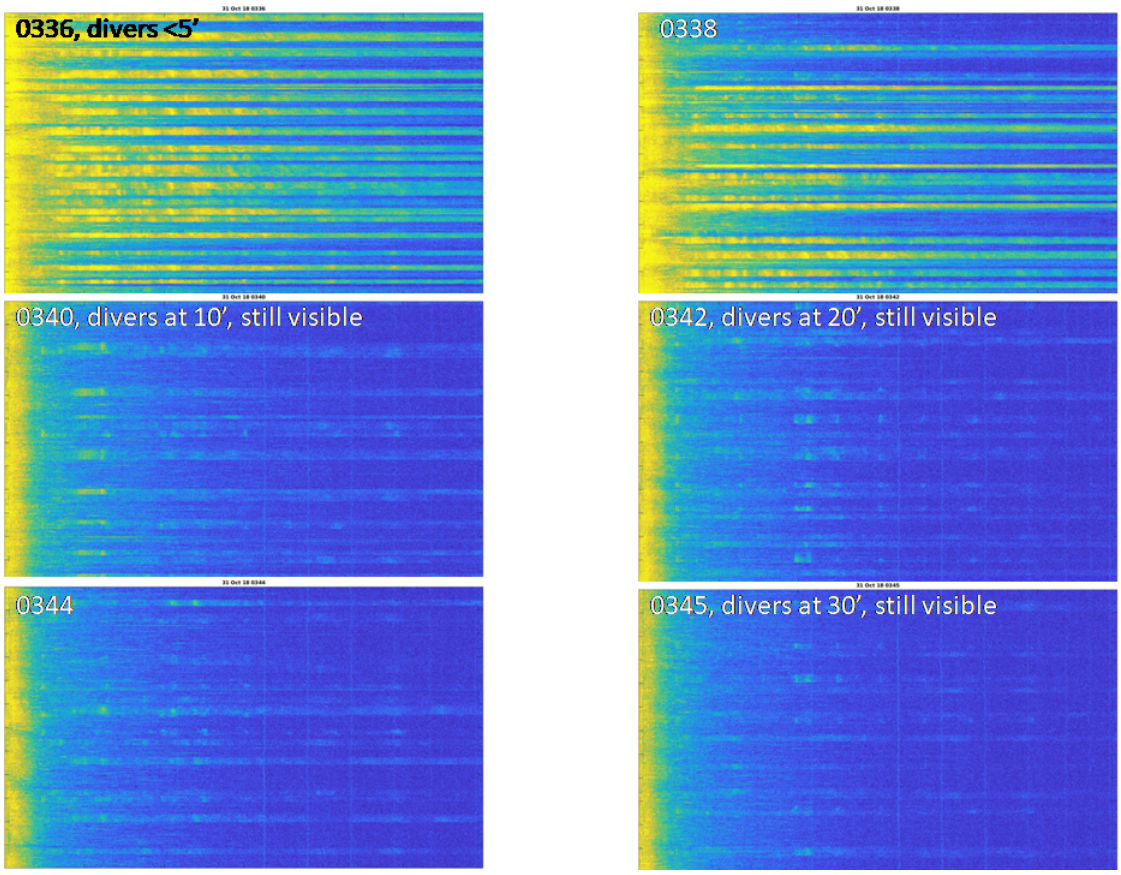


Figure 3-24: Spectrograms of Divers at a Range of 1.5 m to 9.14 m (30 Feet).

from the same dive as the reference signal. Figure 3-25 shows the result of cross-correlation over the same ten minute time interval. It is noteworthy that match filtering appeared to work well while divers were less than 1.5 m from the hydrophone, but the performance quickly deteriorated as diver range increased, lowering to a point where diver transients could not be identified using match filtering. This was likely a due to lower diver signal strength at further ranges.

It is possible that low frequency noise contributed to the poor performance. A high pass, fifth order filter with a cutoff frequency of 5 kHz was applied to all data, including the reference signatures used for match filtering, to remove the low frequency noise. This appeared to improve the performance of match filtering, increasing the detection range to in excess of 12 m. Figure 3-26 - 3-28 depict the results of match filtering using data from the same dive as the control data. In 3-26 divers were less than 1.5 m from the hydrophone, in figure 3-27 they were 3.05 m (10 feet) from the hydrophone, and in figure 3-28 divers

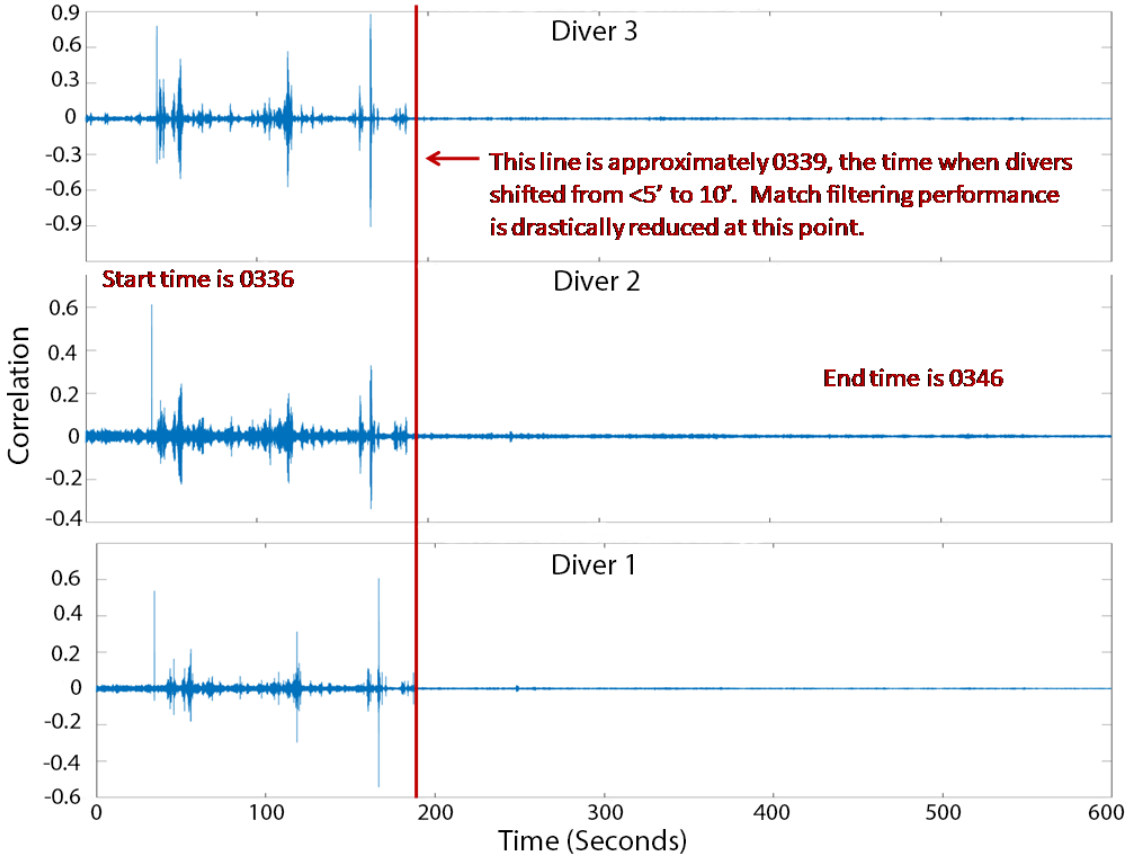


Figure 3-25: Diver Cross Correlation Over 10 Minutes.

were 12.19 m (40 feet) from the hydrophone. Diver transients were clearly visible in the spectrograms for all three of these figures. In figures 3-26 and 3-27 all diver transients that are visible in the spectrogram have corresponding spikes in the match filtering for the same time period. In figure 3-28, the diver transients cannot be seen nearly as clearly as in figures 3-26 or 3-27, but several spikes were still present.

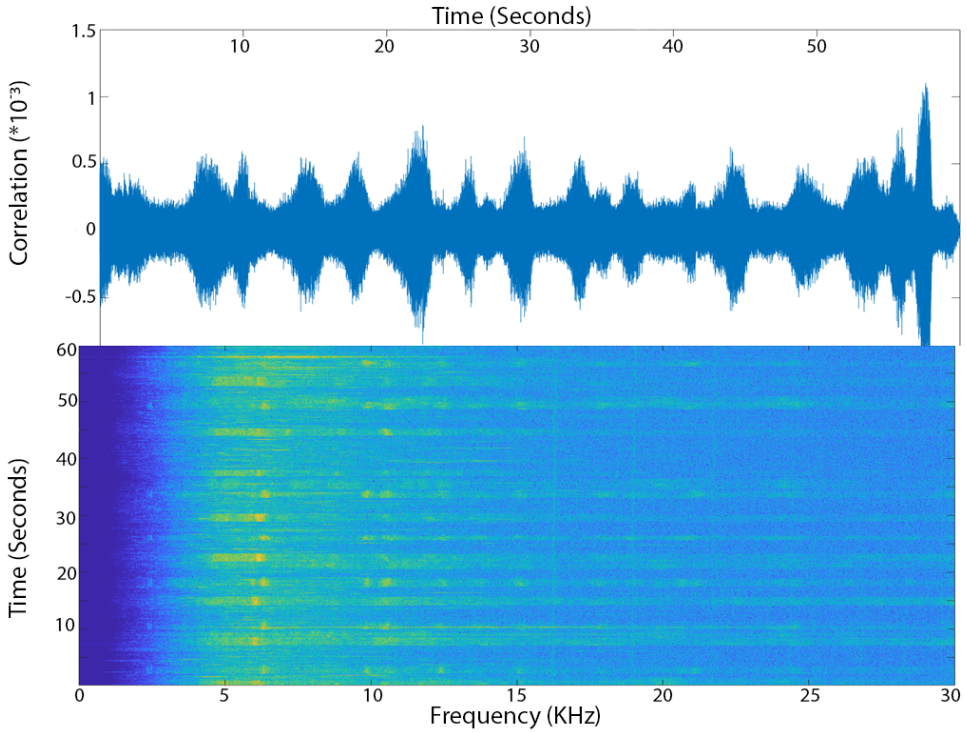


Figure 3-26: Cross Correlation with Divers Range Less Than 1.5 m.

Data from a different dive, 19 October, was compared against the control diver signatures. Again divers were clearly visible in the spectrogram and there were peaks in the cross-correlation that correspond to some of these transients, shown in figure 3-29. Later in the same dive pneumatic hammering commenced as part of the construction work at the Martha's Vineyard ferry terminal. The hammering resulted in roughly 20x higher cross-correlation with the reference signal than the diver transients did. This is shown in figure 3-30. This indicated that the match filtering produced the highest cross-correlation to strong broad band transients, as opposed to the specific signature of divers in the water.

Data where divers were not in the water were analyzed to evaluate how match filtering performed in the presence of broad band transients. Figure 3-31 is from just before divers

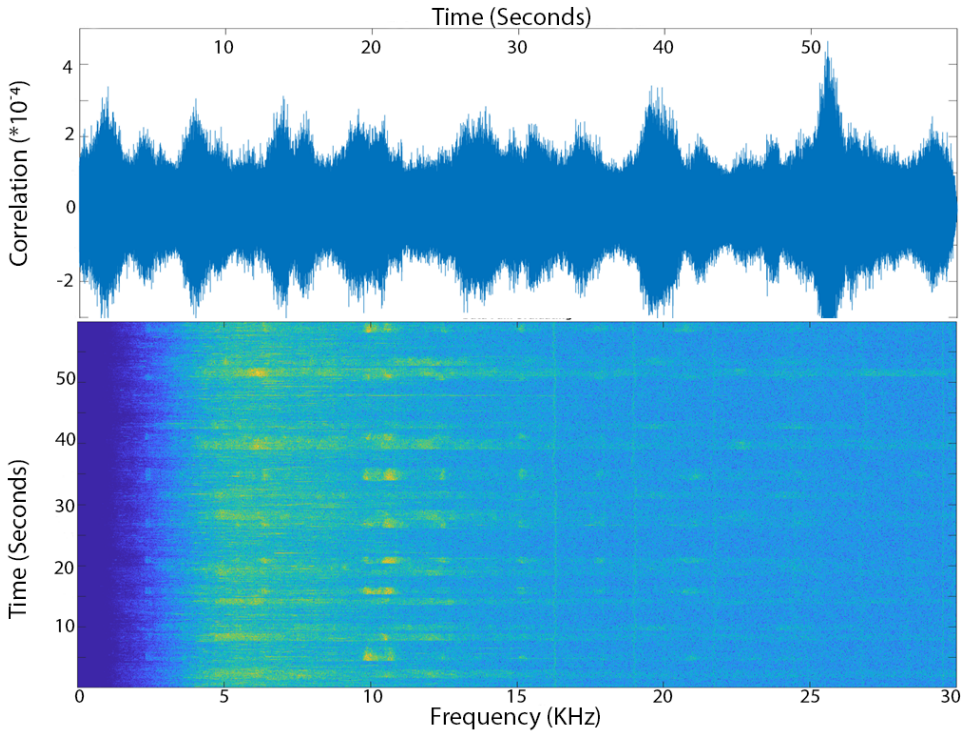


Figure 3-27: Cross Correlation with Divers Range of 3.05 m (10 Feet).

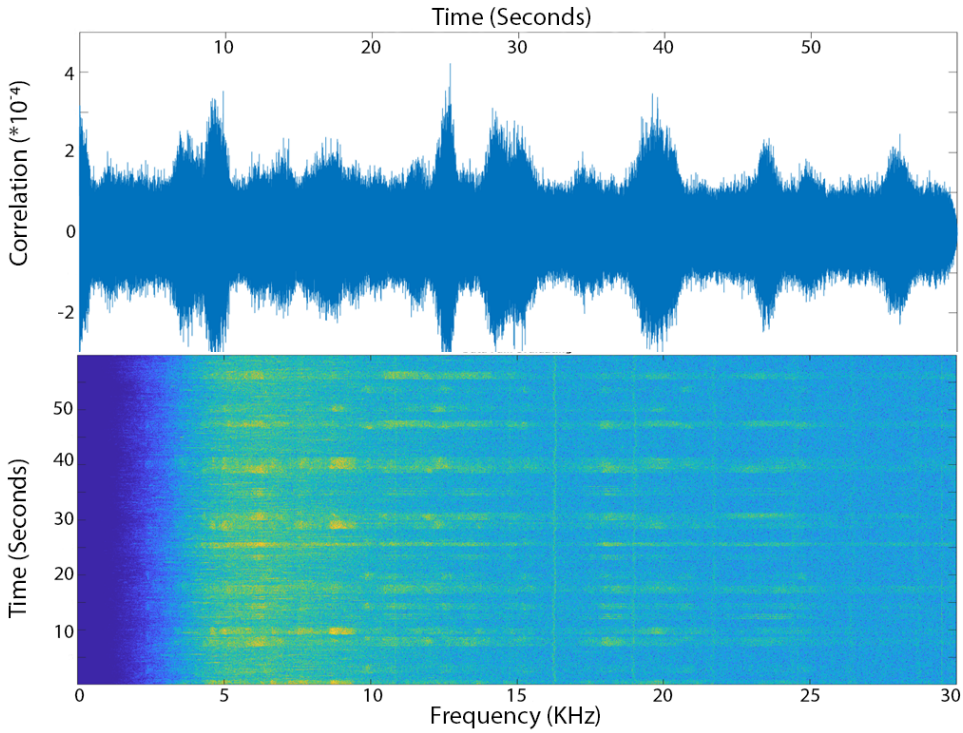


Figure 3-28: Cross Correlation with Divers Range of 12.19 m (40 Feet).

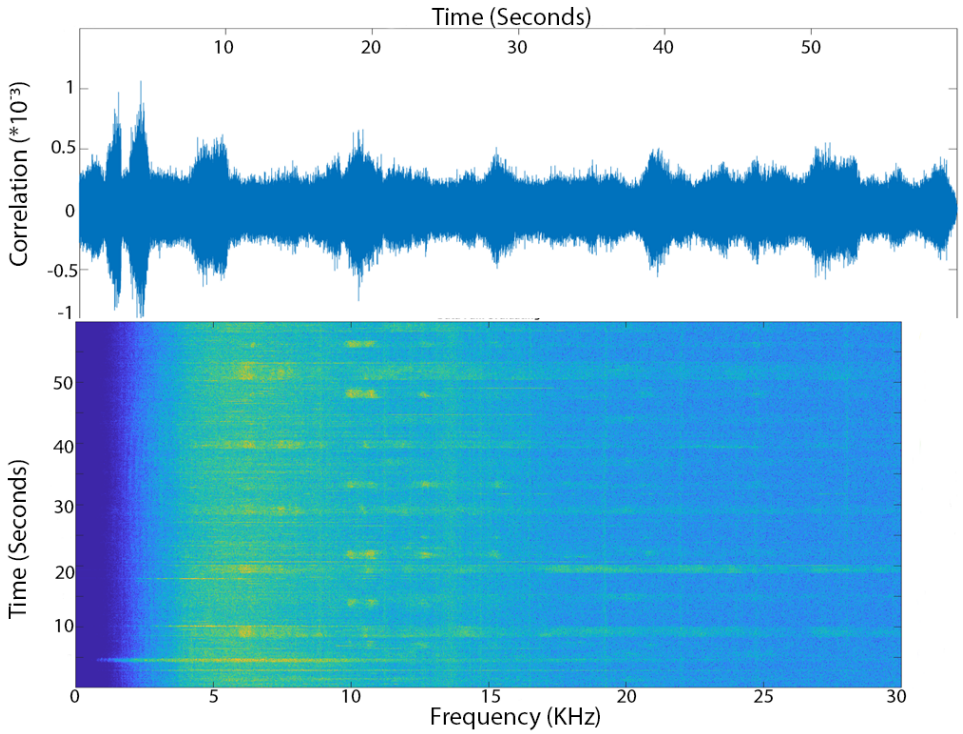


Figure 3-29: Cross Correlation on 19 October 2018 Dive.

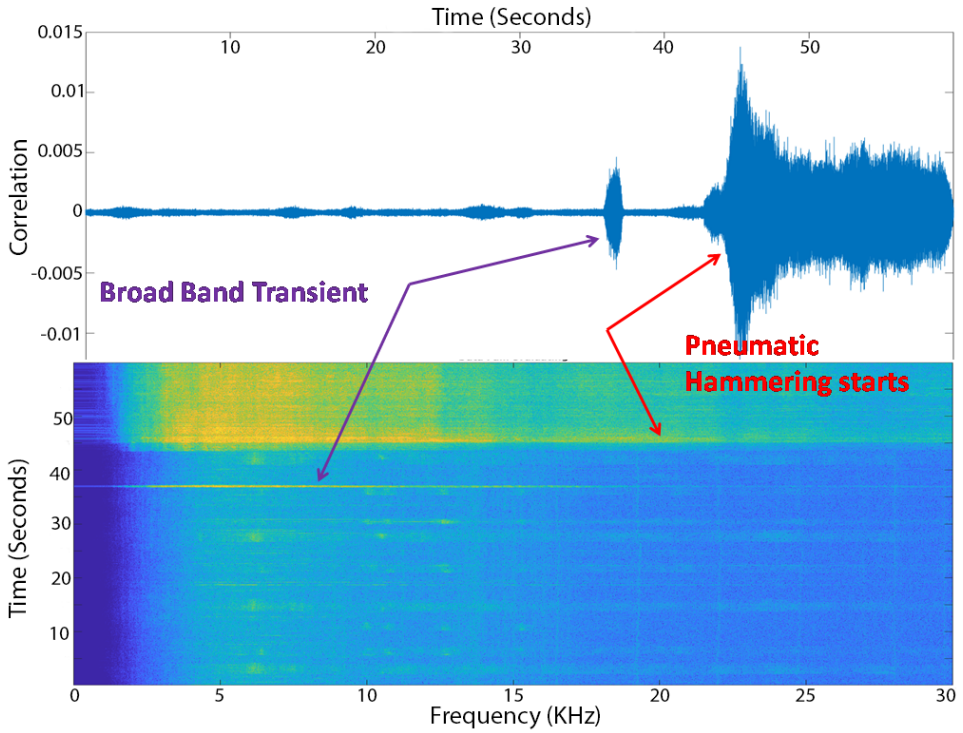


Figure 3-30: Cross Correlation During Construction.

entered the water on 31 October, the same dive as the reference signal. Figure 3-32 is from 31 December, a day when a planned dive was canceled due to safety concerns related to pile driving at the Martha's Vineyard ferry terminal. Both of these figures show large spikes in the cross-correlation corresponding to broad band transients. This indicated that using match filtering for diver detection in an environment where broad band transients were frequently present was not ideal as it would both result in false alarms, shown in figures 3-30 - 3-32, as well as missed detections, shown in figure 3-30. This is because in noisy environments the cross-correlation between reference signal and broadband transients are often multiple orders of magnitude higher than the cross-correlation between the reference signal and other diver breathing transients. The inability to use match filtering techniques directly on the inhale transients of divers was also identified previously by Molbona and Zabarankin [33].

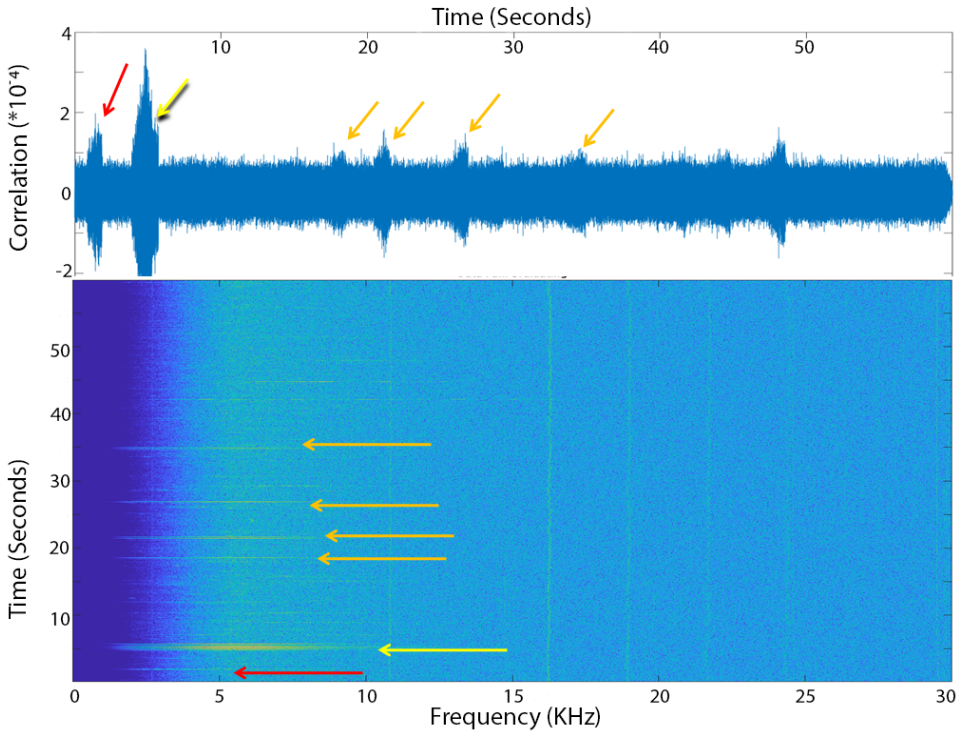


Figure 3-31: Cross Correlation in Presence of Broadband Transients. Arrows Indicate Broadband Transients on Both the Spectrogram and the Cross-Correlation.

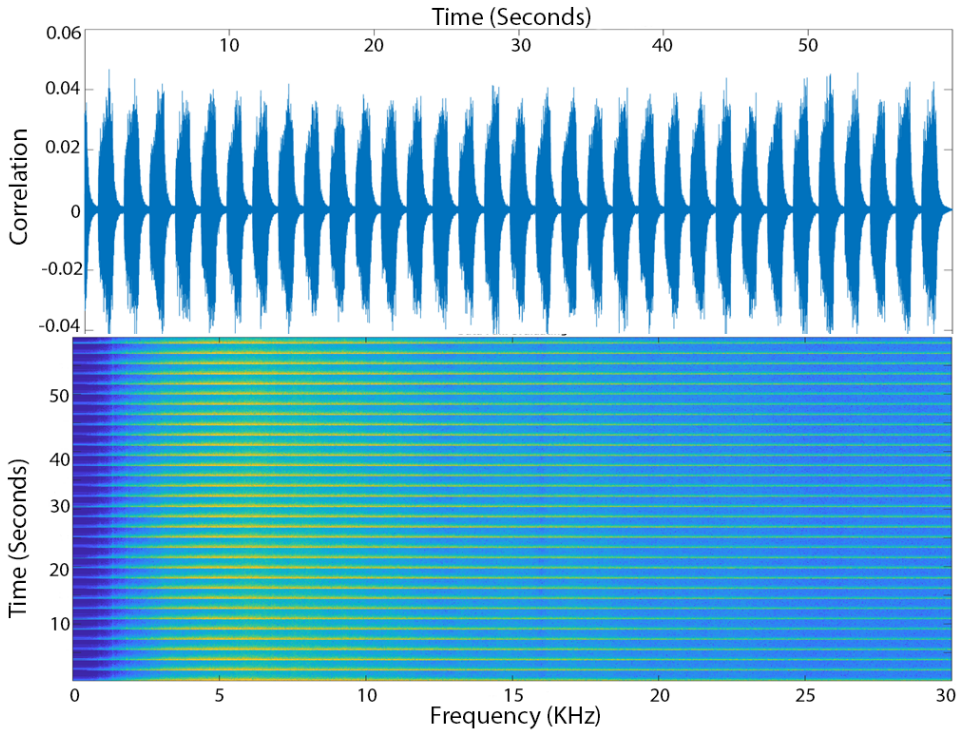


Figure 3-32: Cross Correlation with Pile Driving at the Martha's Vineyard Ferry Terminal.

3.7.2 Match Filtering of Diver Breathing Frequency

An alternative approach to using match filtering for diver detection is to match signals that contain a diver breathing transients over several breathing cycles. Instead of analyzing a single transient, as was attempted in section 3.7.1, multiple transients are used for match filtering. This is performed by taking a Fourier transform on a known signal that contained several diver breathing frequency cycles. This signal is then matched against the Fourier transform of the signal being investigated. If the frequency of transients in both signals matched a high cross-correlation would result and diver detection would be reported. This method for diver detection has been used extensively in the past and is discussed along with its limitations in chapter 2 [51][31][23][42][46]. This type of analysis was not replicated in this thesis due to its extensive previous use.

3.8 Modeling: Propagation Path Evaluation

Physical modeling was used to evaluate the acoustic propagation paths between the divers and the hydrophone. The model was based on acoustic recording of divers at known distances from the hydrophone and the water depth. A total of nine different propagation paths were evaluated during the analysis.

The first step in modeling the acoustic propagation path was data collection. Data from the 0300 hour dive on 31 October was used because it had the lowest background noise. The goal of the acoustic modeling was to determine the acoustic arrival paths from the diver to the hydrophone. This was accomplished by solving for the source level of the divers, determining the absorption coefficient for the bottom, and determining the scattering coefficients of both the surface and the bottom, and then evaluating different propagation paths.

This analysis assumed the bottom was flat and both the diver and the hydrophone were 1 m from the bottom. A constant water depth of 21.366 m (70 feet) was also assumed. These were reasonable assumptions based on the actual conditions of the dives conducted for this thesis. An attenuation constant of 2 dB per km was used, which corresponds to a frequency of 20 kHz at a temperature of 4 C [47].

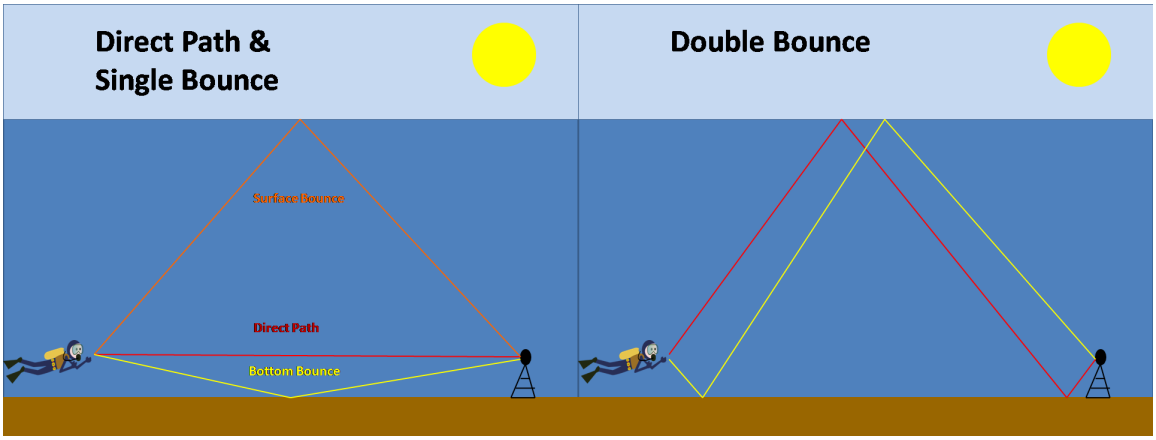
3.8.1 Data Collection

Diver data were evaluated for six distances from less than 1.5 m to 15.24 (50 feet). Multiple diver acoustic emissions were evaluated at each range. The middle one second of data were collected from each transient. These data were high-pass filtered with a cutoff frequency of 5 kHz and an order of 20 to minimize interference by low frequency noise. The absolute values of the acoustic transients for each range were averaged, producing the average acoustic pressure of a diver transient for each range. Three diver transients, one for each diver, were selected for the closest range, as the three divers could be distinguished at this range. Five diver transients were chosen for each of the other five ranges. Five samples of background noise were also taken from the time just before divers entered the water. The average of the RMS background noise was subtracted from the average of the diver signature at each of the ranges, providing the pressure contribution from only the diver transients.

Figure 3-33: Propagation Paths Evaluated

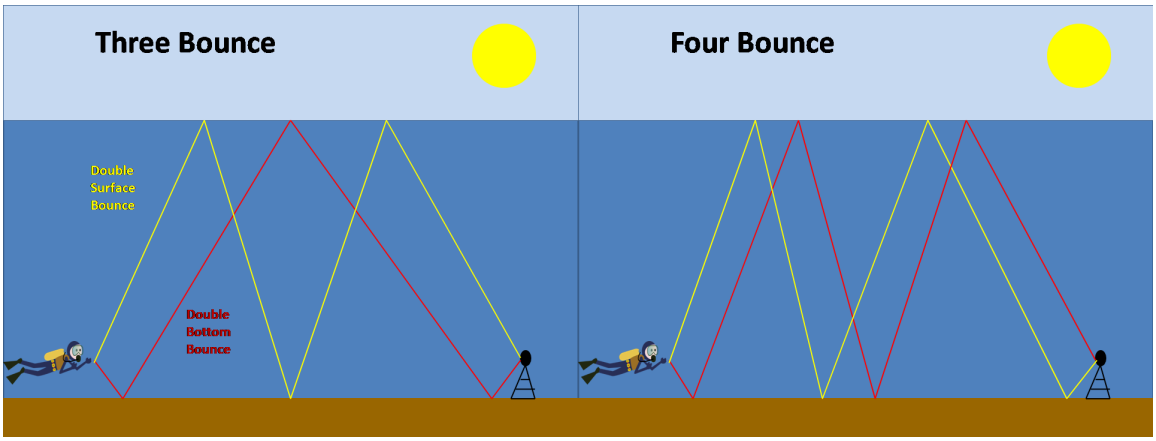
(a) Direct Path and Single Bounce Propagation.

(b) Double Bounce Propagation.



(c) Three Bounce Propagation.

(d) Four Bounce Propagation.



3.8.2 Data Processing

An acoustic model was developed where acoustic pressure at each of the ranges was a function several parameters. The parameters included of the number of arrival paths between the diver and the hydrophone, diver source level, attenuation, absorption by the bottom, and scattering by both the surface and the bottom. Path lengths were calculated for direct path, bottom bounce, and surface bounce, and a combination thereof, for nine separate arrival paths.

$$Path\_Length_{Direct\,Path} = Range_{Diver}$$

$$Path\_Length_{Surface\,Bounce} = \frac{2(Depth_{water} - Altitude_{Diver})}{Sin(Tan^{-1}(\frac{Water_{Depth} - Altitude_{Diver}}{0.5 * Range_{Diver}}))}$$

$$Path\_Length_{Bottom\,Bounce} = \frac{2 * Altitude_{Diver}}{Sin(Tan^{-1}(\frac{Water_{Depth} - Altitude_{Diver}}{0.5 * Range_{Diver}}))}$$

A source sound pressure level was assumed, and later solved numerically. The source pressure was converted to dB and attenuation was subtracted from the source level at a rate of 2 dB per km for all arrival paths. Source level was then converted back to pressure and then reduced to account for spherical spreading for all path lengths. A coefficient for scattering was applied to each surface bounce and a combined scattering and absorption coefficient was applied to each bottom bounce. The RMS pressure of all specified arrival paths was then added to predict the pressure at the specified range from the hydrophone. The number of arrival paths evaluated included all odd numbers from one to nine.

$$L_P\ (dB) = 10log_{10}(\frac{p^2}{p_{ref}^2})$$

$$p_2 = \frac{p_{1m}}{4\pi * Path\_length^2}$$

3.8.3 Model Tuning

The propagation path model was imported into Matlab to solve for source level, the absorption coefficient, the surface scattering coefficient, and a combined bottom bounce coefficient for absorption and scattering. An error term was added to account for measurement error of the background noise. One thousand values for source level between $3.5 * 10^5$ and $1.3 * 10^7$ micro Pascals were tested. The low limit of $3.5 * 10^5$ micro Pascals was chosen because it was the average RMS pressure for divers less than 1.5 m from the hydrophone. The high limit of $1.3 * 10^7$ micro Pascals was selected as this was 3 times the calculated source pressure based

Table 3.11: Acoustic Arrival Path Model Final Parameters.

Number of Arrival Paths	1	3	5	7	9
Mean Squared Error (µPa^2)	$1.9502*10^6$	$1.5429*10^6$	$1.5411*10^6$	$1.5393*10^6$	$1.5391*10^6$
Source Sound Pressure Level (µPa)	$2.3836*10^6$	$1.4256*10^6$	$1.4253*10^6$	$1.4251*10^6$	$1.4249*10^6$
Background Measurement Error (µPa)	$3.7572*10^3$	$3.3340*10^3$	$3.2800*10^3$	$3.2110*10^3$	$3.1824*10^3$
Surface Scattering Coefficient	N/A	1	1	1	1
Combined Bottom Bounce Coefficient	N/A	1	1	1	1

on 1 m range and direct path only propagation. One hundred values, ranging between 0 and 1 for scattering and absorption coefficients, and 0 and $6.4 * 10^3$ micro Pascals for background noise measurement error were evaluated for these parameters.

Every permutation of the values in the preceding paragraph were used to predict an acoustic pressure for a diver transient at each of the 5 ranges between 3.05 and 15.24 m (10 and 50 feet). The model error was determined by subtracting the predicted RMS acoustic pressure from the measured RMS acoustic pressure. This model error was squared and the sum of squared errors for all five distances was added, producing metric used for loss function minimization.

$$Model\_Metric = \Sigma_{Range\ 1:5}(p_{Measured} - p_{Predicted})^2$$

The range of values evaluated for each variable was reduced iteratively for each series of arrival paths analyzed until the mean squared error for loss function reached a minimum. At this point each of the variables was solved. Table 3.11 shows the values of each parameter for a given number of arrival paths. Figure 3-34 depicts the measured acoustic pressure and predicted acoustic pressure for both one and nine arrival paths as a function of range. A discussion of these results is in chapter 4.

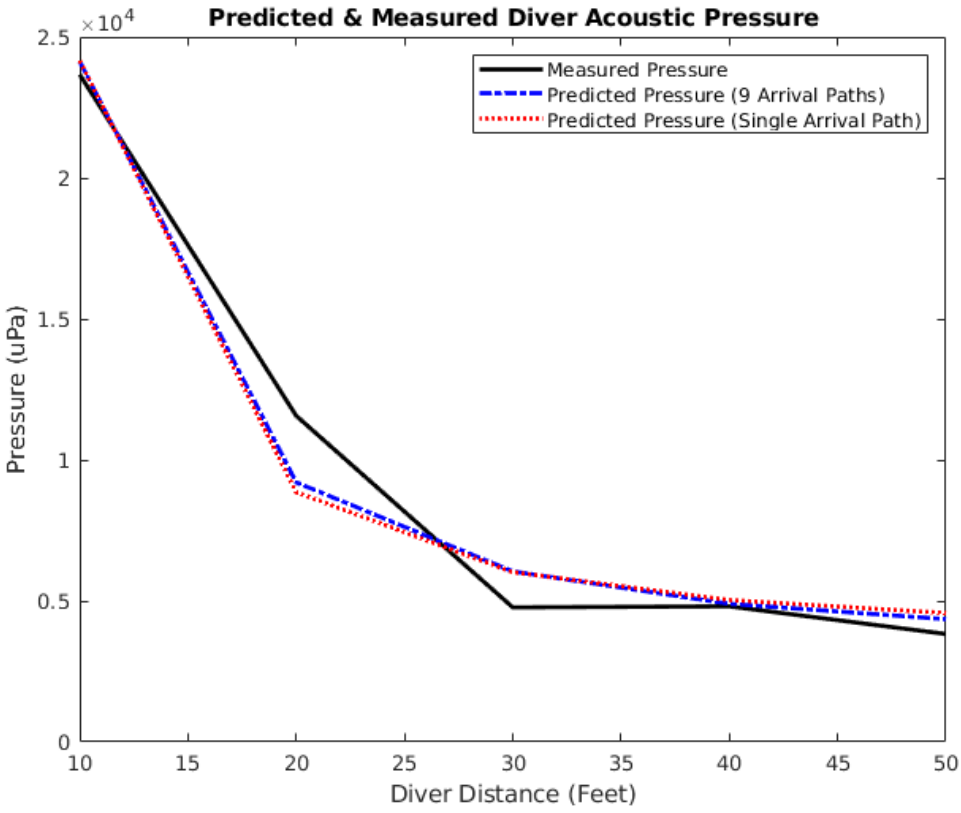


Figure 3-34: Measured Vs Predicted Acoustic Pressure.

Chapter 4

Results and Discussion

This chapter presents the results of this thesis. It begins by examining the machine learning model's performance against previously unseen data including from dives not used in model development and validation data specifically set aside for this model evaluation. The ability of the model to detect divers during short breaks in high ambient noise is then evaluated. The model performance with sequential data processing is discussed next, and the chapter is concluded by examining the results of the acoustic modeling presented in chapter 3.

4.1 Machine Learning Models

In the course of this thesis 84 models were trained to eventually produce 2 fully tuned machine learning models. The model training and tuning process is presented in section 3.6.5. The single frequency model was trained and tuned using an early data set that consisted of data from the 8-15 kHz band only. When new data were produced through subsequent dives, a large fraction was added to the training set to produce an iteratively refined model.

The dual frequency machine learning model was trained and tuned with data from all 15 dives through 9 January 2019. In addition a second frequency band of 18-25 kHz was added, nearly doubling the data set. These data were used to produce the final machine learning model in this thesis. Four additional dives were conducted after 9 January 2019. These dives were used to validate the performance of the final machine learning model and were not used to produce a new model. Figure 4-1 shows the data used to train and validate the models presented in this section.

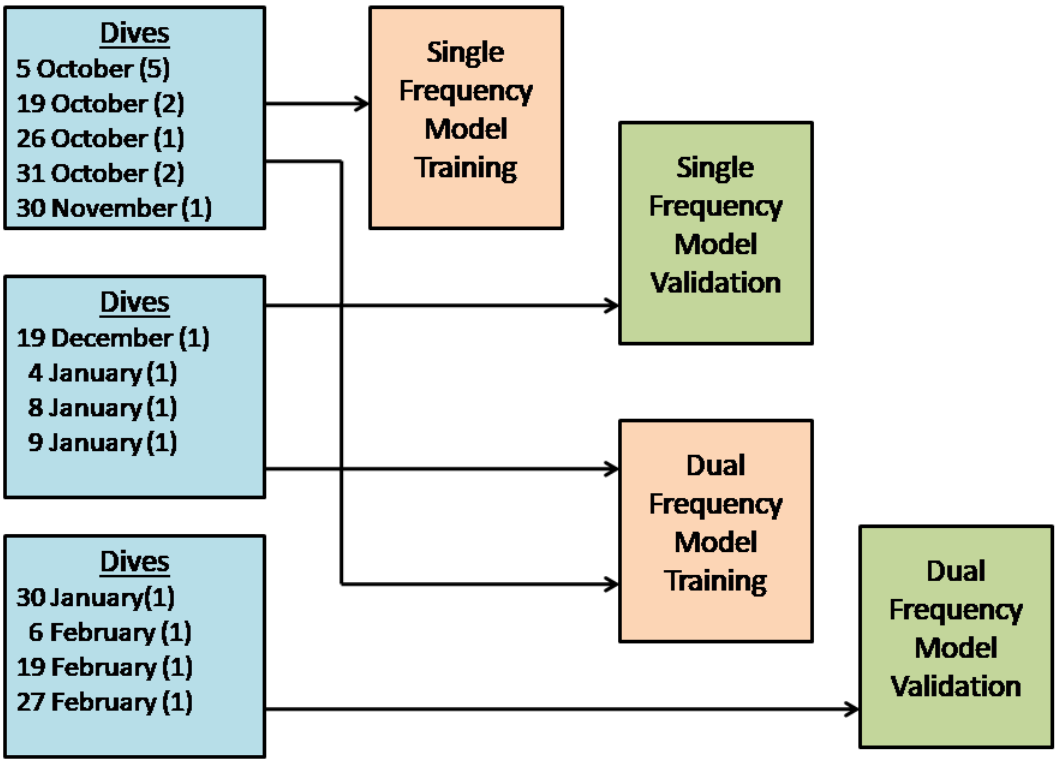


Figure 4-1: Model Generation and Independent Validation Data.

4.2 Model Evaluation with New Data

This section evaluates the performance of both the single frequency and dual frequency models on data collected after model generation. The models' performance is shown for both frequency bands for each dive. A combined confusion matrix depicting the models' overall performance on each dive is also shown.

4.2.1 Single Frequency Model Performance on New Data

Data from the 19 December 2018, 04 January 2019, 08 January 2019, and 09 January 2019 were tested using the previously trained single frequency model. The single frequency model used to evaluate these data was produced with data up to and including 30 November dive. This model was trained using only the 8-15 kHz band but was tested against both frequency bands. Testing of the 4 additional dives resulted in the model classifying 92.6% of the spectrograms correctly, producing an average weighted F1 score of 0.925. Table 4.1 shows models performance on each frequency band for the subsequent dives. Figure 4-2 is the combined confusion matrix for all of the data sets in table 4.1.

Table 4.1: Single Frequency Model Performance on New Data.

Data Set	Weighted Average F1 Score	Divers F1 Score	No Divers F1 Score	Support
19 Dec 8-15 kHz	0.98	0.98	0.98	42
19 Dec 18-25 kHz	0.84	0.84	0.84	50
4 Jan 8-15 kHz	0.9	0.91	0.9	62
4 Jan 18-25 kHz	0.85	0.87	0.82	140
8 Jan 8-15 kHz	0.95	0.95	0.95	202
8 Jan 18-25 kHz	0.99	0.99	0.99	178
9 Jan 8-15 kHz	0.95	0.95	0.95	238
9 Jan 18-25 kHz	0.89	0.9	0.88	246

During the tuning process the model produced an F1 score of 0.87. It is notable that the new data performed better than the testing data used to tune it. This was likely because the new data were disproportionally at a closer range than the testing data and therefore contained more discernible features of divers. This shows that a convolutional neural network can be used to evaluate new data and classify it correctly the majority of the time.

There has been no previously published work regarding scuba diver detection with a convolutional neural network; therefore, there is no baseline metric for success. A recently

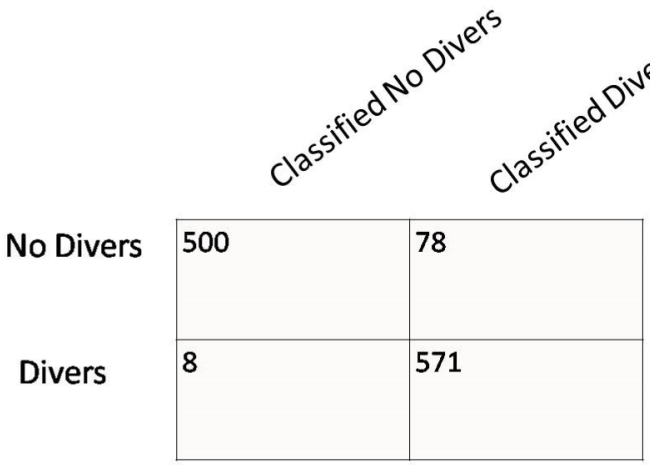


Figure 4-2: Single Frequency Model Combined Confusion Matrix for 19 December 2018 - 9 January 2019 Dives.

published review paper regarding medical image analysis with convolutional neural networks evaluated contemporary convolutional neural networks for medical image classification. The 10 networks analyzed had an accuracy ranging from 75% to 99.8% with an average accuracy of 88.4% [6]. With an accuracy of 92.6% the convolutional neural network used in this thesis performed at a level comparable to other modern convolutional neural networks.

4.2.2 Dual Frequency Model Performance on New Data

The dual frequency model was trained and tuned with data through the 9 January 2019 dive and was used to evaluate all subsequent data. The new model classified 91.3% of the spectrograms from the later dives correctly, resulting in a combined average F1 score of 0.910. Table 4.2 shows the model's performance for both frequency bands for following dives. Figure 4-3 is the combined confusion matrix for all of the data sets presented in table 4.2.

4.2.3 Explanation of Outliers

The model generally performed well on new data with exceptional performance on 30 January and 6 February, modest results on 27 February, and relatively poor performance on 19 February. Dive specific performance can be explained in part by the conditions of the dives.

The performance on 30 January and 6 February 2019 is likely a result of being conducted

Table 4.2: Dual Frequency Model Performance on New Data.

Data Set	Weighted Average F1 Score	Divers F1 Score	No Divers F1 Score	Support
30 Jan 8-15 kHz	0.91	0.92	0.91	236
30 Jan 18-25 kHz	0.94	0.95	0.99	248
6 Feb 8-15 kHz	0.99	0.99	0.99	266
6 Feb 18-25 kHz	0.98	0.98	0.98	244
19 Feb 8-15 kHz	0.73	0.7	0.76	110
19 Feb 18-25 kHz	0.8	0.77	0.84	78
27 Feb 8-15 kHz	0.83	0.84	0.82	216
27 Feb 18-25 kHz	0.91	0.92	0.91	252

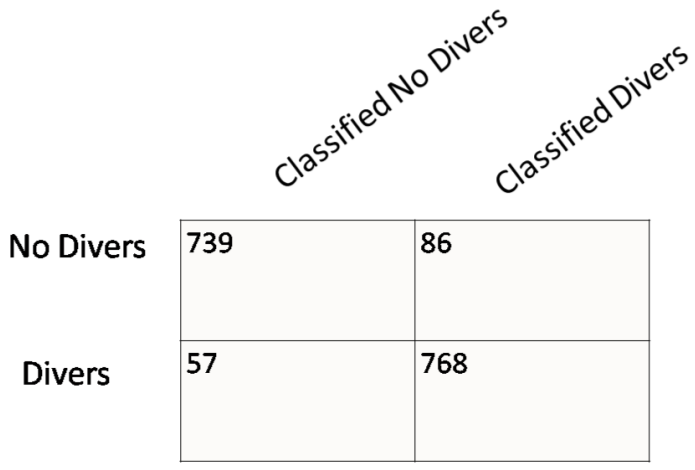


Figure 4-3: Dual Frequency Model Combined Confusion Matrix for 30 January - 27 February 2019 Dives.

during the 0500 hour with significantly less background noise than other dives. The 19 February 2019 dive did not follow the normal testing protocol but instead divers surveyed the WHOI dock and were in excess of 15 m from the hydrophone for the majority of the dive. This dive also occurred during a period of heavy construction at the Martha's Vineyard ferry terminal. These two factors were likely responsible for the comparatively poor performance of the model on this dives. The overall performance of the machine learning, classifying 91.3% of the spectrograms from new dives correctly, confirms the validity of using low cost passive sonar and a convolutional neural network for open circuit diver detection.

4.3 Model Evaluation with Validation Data

As discussed in section 3.6.1 the machine learning images were split into three groups, training, testing, and validation prior to training a model. All data through the 9 January 2019 dive was divided into these groups. Data from the subsequent dives was not split because it was used for model validation only. The training data were used to train the model and the testing data were used to tune the model. The validation data were set aside to independently evaluate the model's performance. The machine learning model classified 93.4% of the validation spectrograms correctly, producing a weighted average F1 score of 0.93. The results of this analysis are shown in table 4.3 and figure 4-4.

Table 4.3: Dual Frequency Model Performance on Validation Data.

Data Set	Weighted Average F1 Score	Divers F1 Score	No Divers F1 Score	Support
Validation Data	0.93	0.94	0.93	304

The validation data outperformed the training data when evaluated by the dual frequency model. The validation data had an F1 score of 0.93 compared to 0.87 for the training data. This suggests that over-fitting did not occur during the training process and that the model generalized to new data well. The fact that the validation data outperformed the training data suggests that a comparatively higher portion of the validation set had features that the model was able to easily distinguish. This, along with evaluating data from dives that were not involved in the training or tuning of the model, indicated that the model performed well, regardless of the data that it was provided.

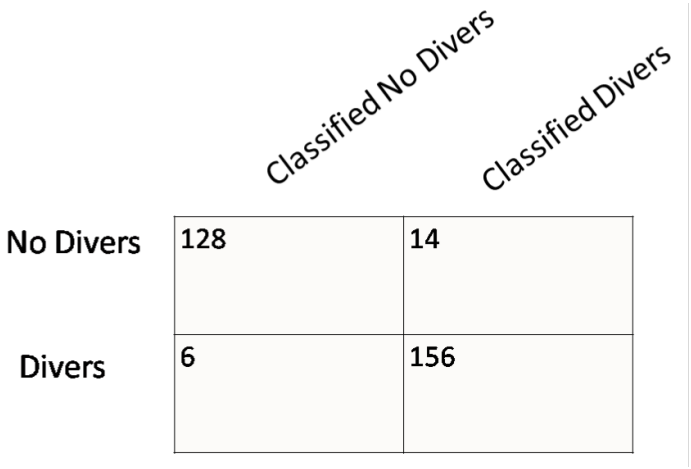


Figure 4-4: Dual Frequency Model Confusion Matrix for Validation Data.

4.4 Diver Detection Through Noise

The dive on 18 January 2019 was used to evaluate the machine learning model's ability to identify divers in brief breaks in high ambient noise. The background noise was caused by a tug boat participating in construction at the Martha's Vineyard ferry terminal. There were three lulls in the ambient noise, which contained a total of four diver transients detectable by the author. The lengths of the breaks were 4 seconds, 9 seconds, and 15 seconds. The machine learning model properly classified three of the four transients correctly. The misclassified transient was during the same break in the noise as one of the properly classified transients. This demonstrated that the model was able to detect divers based off a single acoustic transient and that divers could be detected during brief cessations in background noise. Figures 4-5 - 4-7 show the one minute spectrograms that contain the break in the noise and the diver transients with their classification results.

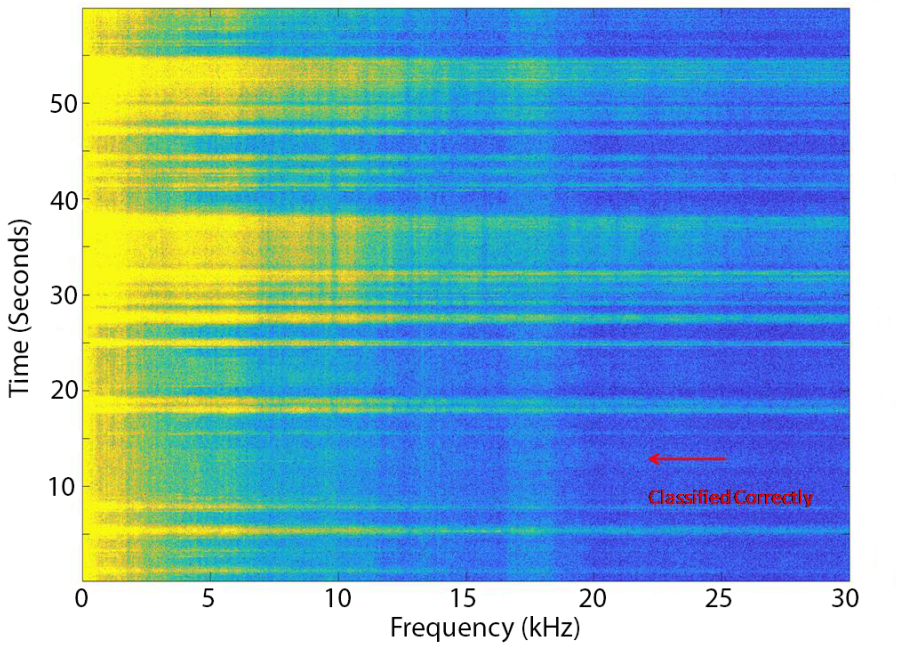


Figure 4-5: Diver Detection Through Noise 1.

The fact that the model was able to properly classify three out of four diver transients and detect divers in all three of the breaks in background noise showed the flexibility of convolutional neural networks for diver detection. Traditional diver detection requires upwards of 10 diver breathing cycles and as a result would not have been able to detect divers in this case [46]. This indicates that convolutional neural networks have the potential to be more

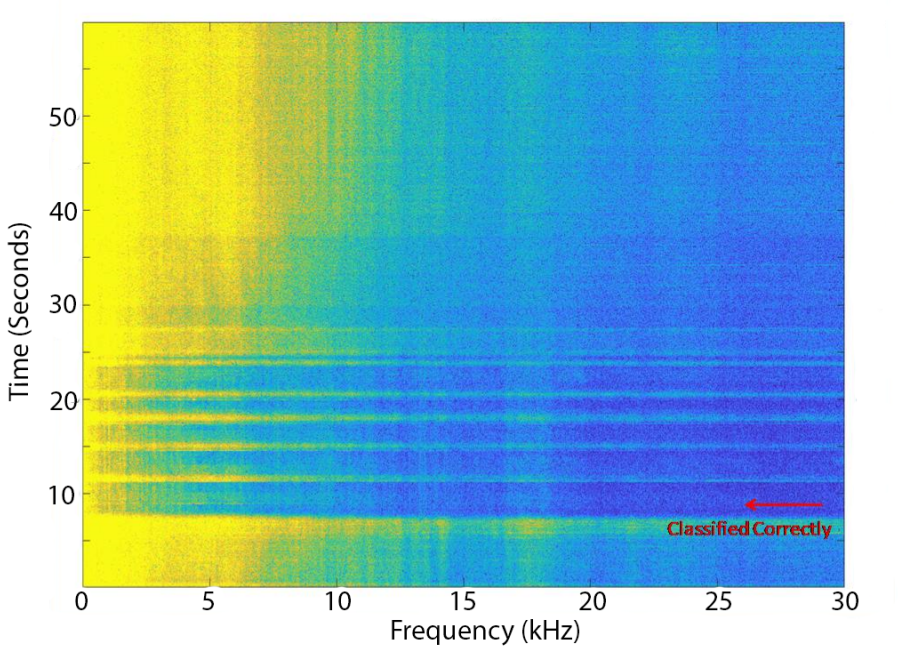


Figure 4-6: Diver Detection Through Noise 2.

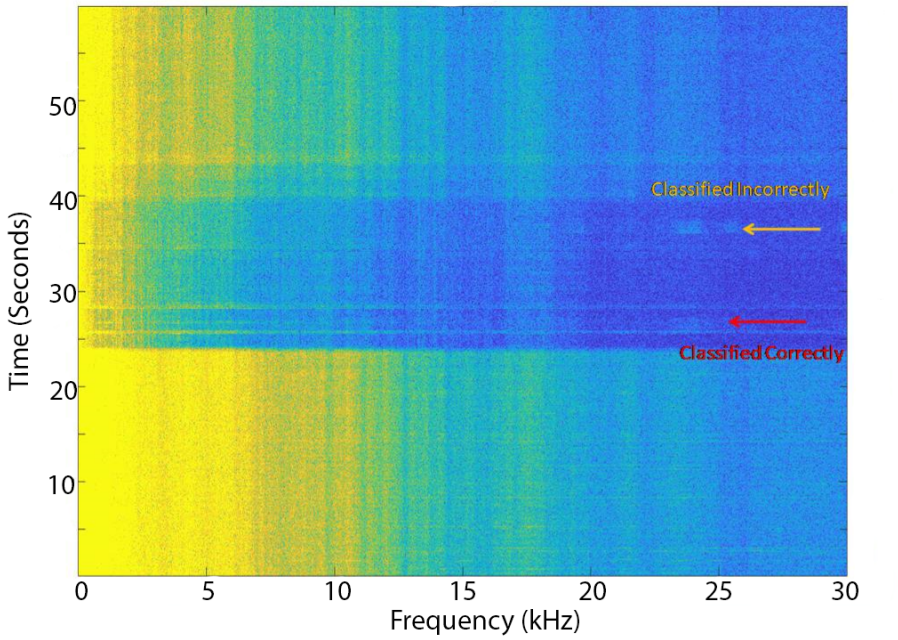


Figure 4-7: Diver Detection Through Noise 3.

versatile for diver detection than traditional diver detection methods [51][31][23][42][46]. Figure 4-8 shows the construction at the Martha's Vineyard ferry terminal. This image was taken from the WHOI pier and shows the tug that produced the majority of the background noise on 18 January.

Figure 4-8: Construction and Tug Boat at Marta's Vineyard Ferry Terminal 18 January 2019.

4.5 Diver Detection with Sequential Processing

An evaluation was conducted to determine if processing data sequentially could eliminate false positives and missed detections by including previous results in the classification decision. Sequential processing, described in section 3.6.6, was used to evaluate data from four dives and related periods with no divers. For the time periods evaluated 31 detections were missed by the dual frequency machine learning model where a human was able to identify divers and there were 65 false positives. Sequential processing filtered 20 of the missed detections and 40 of the false positives. The result was that 64.5% of the missed detections and 61.5% of the false positives were removed. Prior to sequential processing the machine learning model correctly classified 93.0% of the spectrograms. With sequen-

tial processing the machine learning model achieved a 97.1% accuracy demonstrating that sequential processing improved the machine learning model's performance.

Low confidence diver detection was defined as a single independent detection. The advantage of low confidence detection was an integration time of 10 seconds, the same as parallel processing. This came at the expense of removing to ability to filter false positives, as every time the model classified a spectrogram as divers, diver detection was reported. The Markov chain model for low confidence diver detection is shown in chapter 3 in figure 3-22. Table 4.4 shows the required Markov state for diver detection as a function of confidence level.

Table 4.4: Diver Detection Prediction as a Function of Markov State and Confidence Level.

	Markov Model State						
	State 0	State 1	State 2	State 3	State 4	State 5	State 6
Low Confidance	No Divers	Divers	Divers	Divers	Divers	Divers	Divers
Medium Confidance	No Divers	No Divers	Divers	Divers	Divers	Divers	Divers
High Confidance	No Divers	No Divers	No Divers	Divers	Divers	Divers	Divers

Medium confidence diver detection required two sequential detections. This came at the cost of an additional 10 seconds of integration time over low confidence detection. Raising the classification threshold to medium had the advantage of filtering out single spurious diver detections. However, choosing this classification threshold would likely precluded diver detections during brief breaks in background noise, like those shown in the previous section. The Markov chain model for medium confidence diver detection was shown chapter 3 in figure 3-20.

High confidence diver detection required the Markov chain to reach state three before indicating diver detection. This raised the required integration time to at least 30 seconds but lowered the probability of false positives. High confidence diver detection further lowered the likelihood of detecting divers through brief breaks in background noise. The Markov chain model for high confidence diver detection was shown chapter 3 in figure 3-23.

Medium confidence was chosen for diver detection to minimize the required integration time while filtering single spurious false positives. If the penalty for missed detections was high compared to that of false positives the confidence level would have been set to low. Conversely if the cost of spurious alarms was high compared to the price of missed detection, the threshold would have been set to high.

An alternative to changing the confidence level is the use of an asymmetric counter. If the penalty for false alarms was high, the counter would increment at a value less than one and decrement with a value of one. If the penalty for missed detections was high the counter would increment at a value greater than one and again decrement with a value of one. Incrementing at a value of two thirds, with a medium confidence detection confidence is equivalent to using a high confidence detection threshold. Incrementing with a value of two with a medium confidence detection threshold is equivalent to using low confidence as the detection threshold.

Figures 4-9 and 4-10 show the results of sequential data processing for the time just preceding the dive and the dive on 30 January 2019 respectively. These figures depict both the confidence level, one being low confidence, two, medium confidence, and three, high confidence, along with the overall state of the Markov chain. Figure 4-9 shows that there were no false positives via manual classification but that the machine learning model produced four. With a medium confidence classification threshold, none of the false positives resulted in diver classification. As a result the machine learning model, using sequential processing, performed on par with human classification.

Figure 4-10 identifies that human classification failed to identify divers in four 10 second intervals where divers were present. The machine learning model failed to identify divers in seven periods, two sequentially, when divers were present. It is noteworthy that the machine learning model correctly identified divers in the 10 second window after the last positive human classification. With a medium confidence classification threshold, the machine learning model with sequential processing performed at least as good as human classification. This performance demonstrated the potential for machine learning to perform as well as a human under certain circumstances.

Figures 4-11 and 4-12 are similar to those in figures 4-9 and 4-10 and illustrate the same basic results; however, figures 4-11 and 4-12 are from 6 February 2019. It is noteworthy that there was minimal background noise during this dive and therefore the machine learning model performed very well, correctly classifying 98.6% of the spectrograms classified correctly by a human before sequential processing. Figure 4-11 shows that the machine learning model identified divers in two spectrograms during a time that divers were not in the water. The human evaluation of the data at the same time did not produce any false positives.

With a medium confidence threshold for diver detection, no diver detections occurred

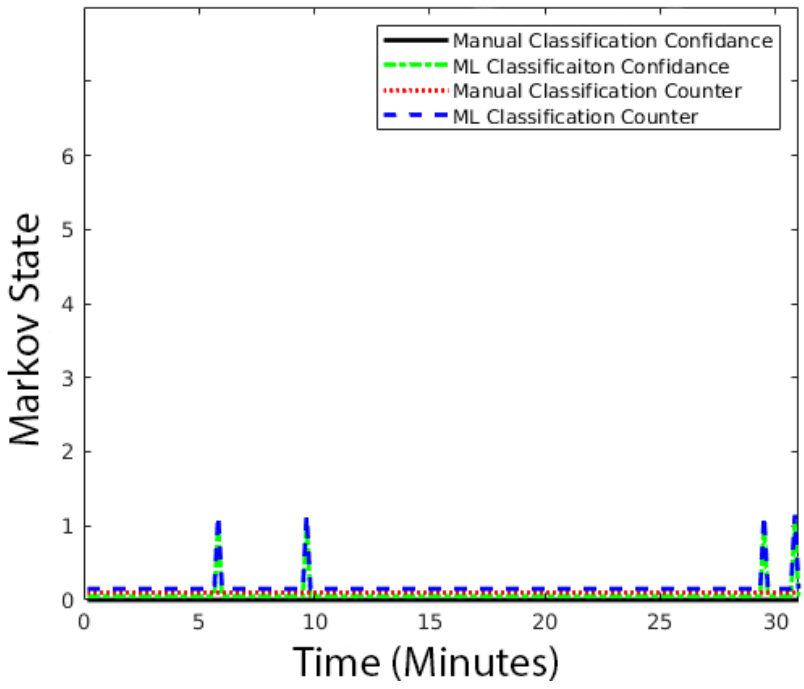


Figure 4-9: Sequential Data Processing, Divers not Present, 30 January 2019.

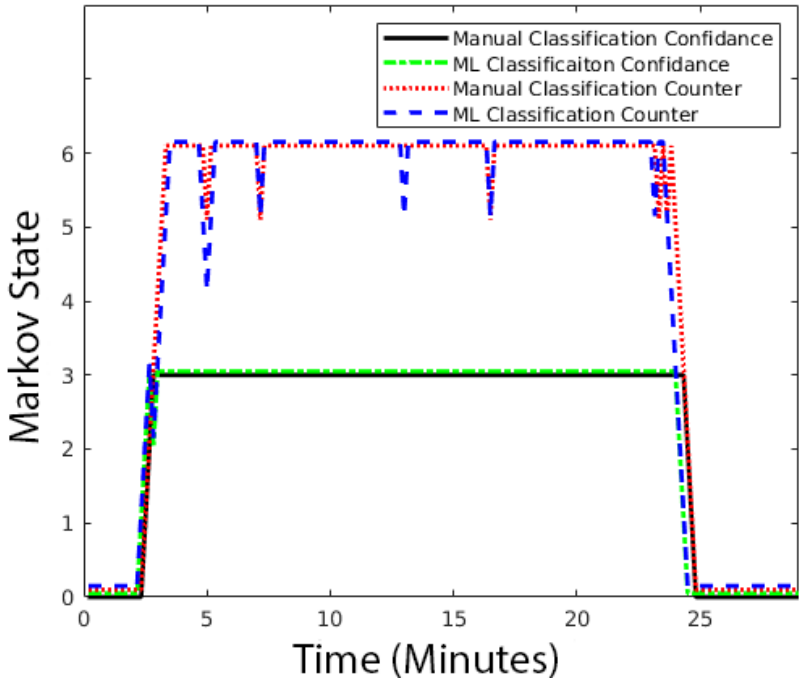


Figure 4-10: Sequential Data Processing, Divers Present, 30 January 2019.

during this time interval for either the human or machine learning model. Figure 4-12 shows three missed detections, two of which corresponded with missed detections of the human evaluator. Additionally, the machine learning model reported a detection significantly after all other detections indicating that it was likely a false positive as opposed to a valid detection. Neither the false positive nor three missed detections changed the overall classification. It is important to note that the human evaluator identified divers in three 10 second intervals prior to the first machine learning diver detection; therefore, the human slightly outperformed the machine learning model during this time period.

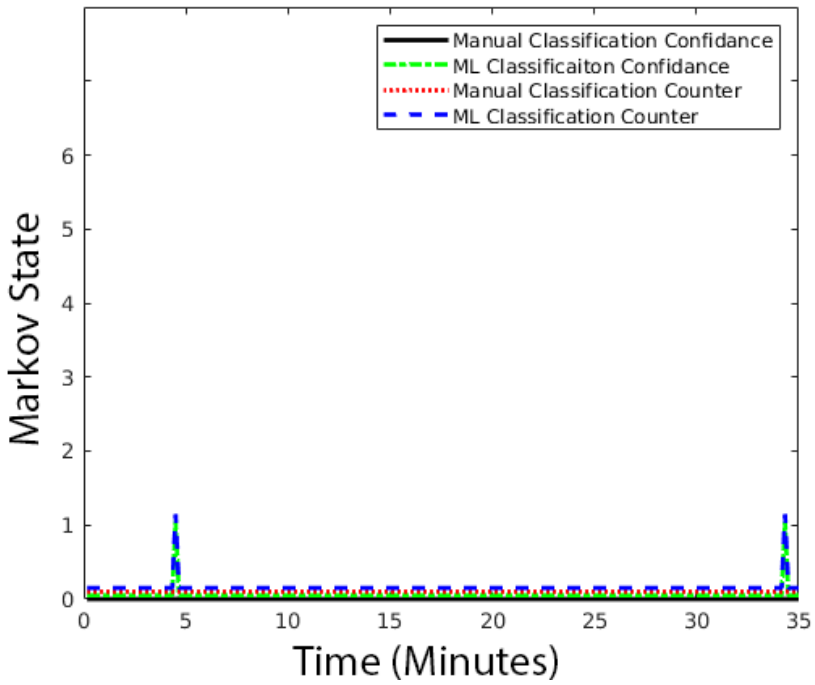


Figure 4-11: Sequential Data Processing, Divers not Present, 06 February 2019.

Figures 4-13 and 4-14 are also similar to figures 4-9 and 4-10, but are from 19 February 2019. Heavy marine construction was ongoing before and during this dive. Additionally, this dive did not follow the normal testing protocol, instead divers surveyed the WHOI pier structure, which resulted in the majority of the dive occurring at a distance in excess of 15 m from the hydrophone. These two factors likely contributed to the poor performance of the model in comparison to the previous two dives.

Figure 4-13 shows that there were several false positives during the 31-minute period

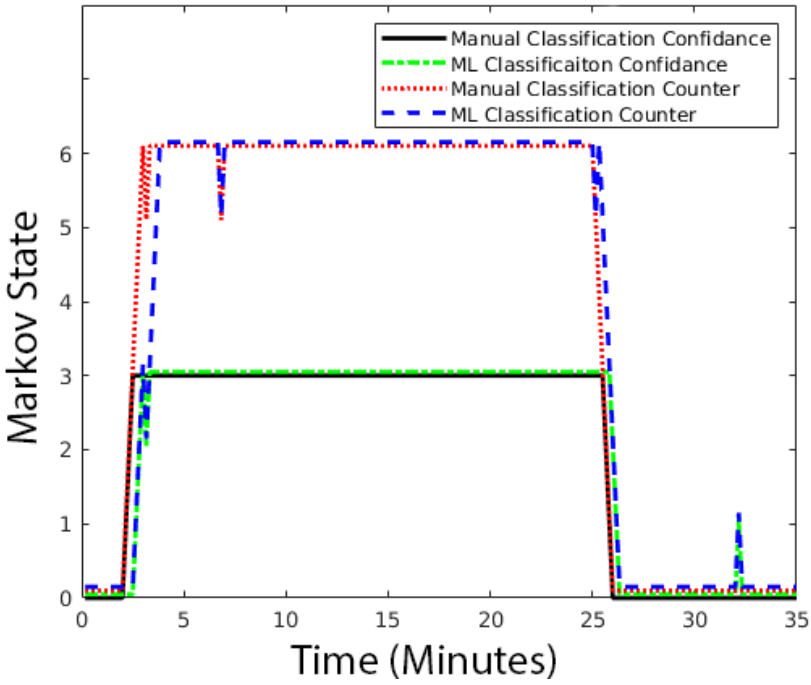


Figure 4-12: Sequential Data Processing, Divers Present, 06 February 2019.

leading up to the dive. Sequential processing was successful at filtering out 13 of the false positives, however consecutive, or near consecutive false positives resulted in the classification of divers when no divers were in the water on 8 occasions. Figures 4-15a - 4-15c show three sequential 10 second spectrograms classified as containing divers during a period when divers were not in the water. Instead they depict pile driving at the Martha's Vineyard ferry terminal and occurred in the 20th and 21st minute of figure 4-13. Manual evaluation of these spectrograms, as well as others resulting in false positives indicated that the machine learning model was likely to produce false positives when there was an abrupt, broadband transition from relative noise to relative quiet, with both periods of noise and quiet in excess of one second.

Figure 4-14 shows that the machine learning model processed sequentially, with a classification threshold of medium confidence, performed nearly as well as the human evaluation for this same time period. It is noteworthy that the human evaluator, the author, was able to detect divers one minute before the machine learning model. Additionally, sequential processing filtered out several valid diver detections because they were not consecutive.

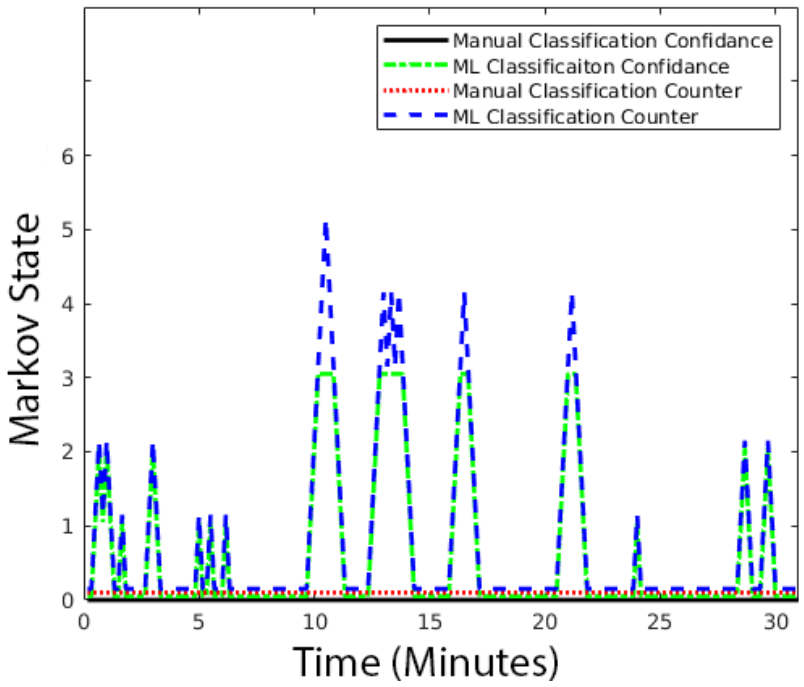


Figure 4-13: Sequential Data Processing, Divers not Present, 19 February 2019.

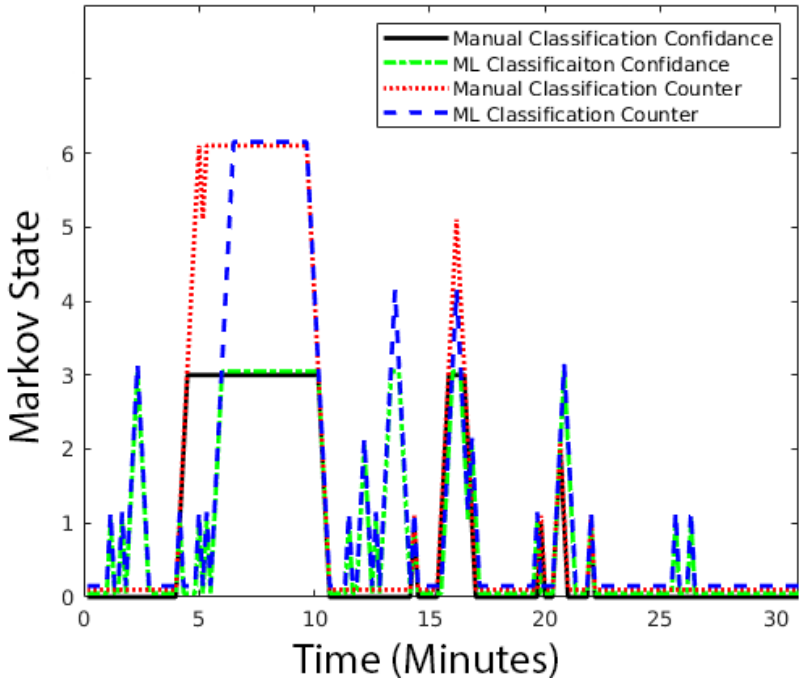
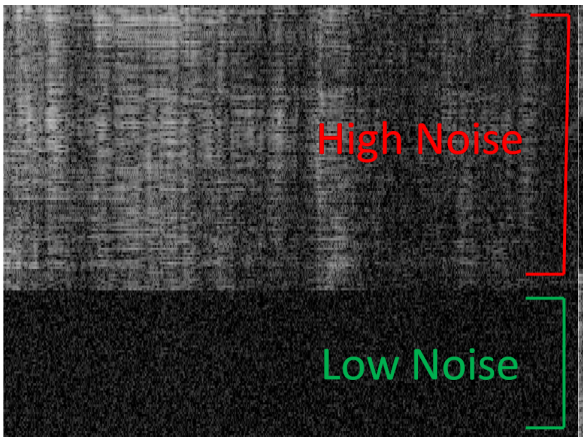


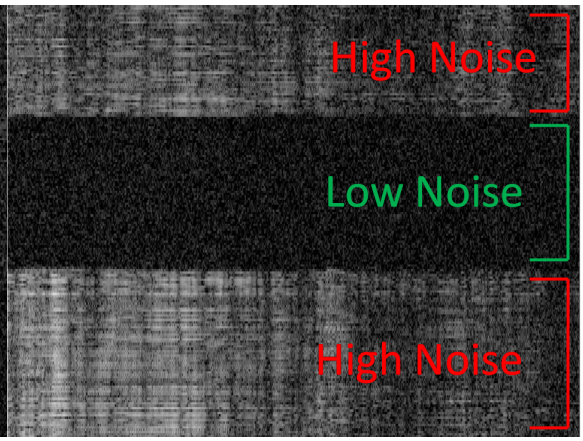
Figure 4-14: Sequential Data Processing, Divers Present, 19 February 2019.

Figure 4-15: Ten Second Machine Learning Spectrograms Producing False Positives Examples. False Positives Frequently Occurred During Abrupt, Broadband Transition from Relative Noise to Relative Quiet, with Both Periods of Noise and Quiet in Excess of One Second.

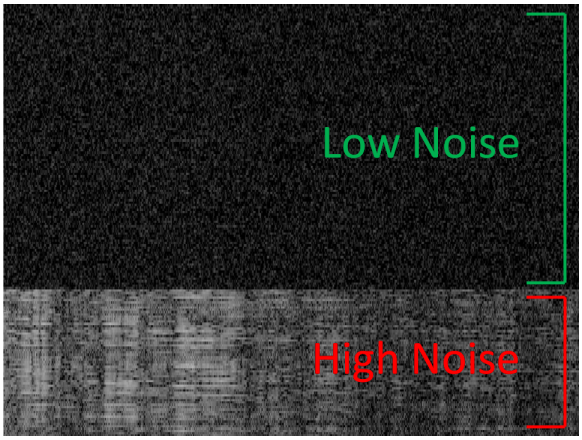
(a) Example 1.



(b) Example 2.



(c) Example 3.



The twentieth and final dive for this thesis occurred in the 0700 hour of 27 February 2019. Figure 4-16 shows the result of sequential processing for a period of 35 minutes prior to the dive and figure 4-17 shows the result while divers were in the water. It is important to note that during these times light construction activity at the Martha's Vineyard ferry terminal was ongoing and several vessels passed near the hydrophone. As a result, the performance was modest compared to the 30 January or 6 February 2019 dives but better than the 19 February dive. During the period when divers were not in the water there were 28 false positives. Using a medium confidence detection threshold, sequential processing was successful at filtering out 21 of the false positives, however divers were incorrectly identified on 5 occasions. When divers were in the water, the machine learning model, along with sequential processing properly identified divers two minutes before a trained human operator. Once the divers were detected by both the machine learning model and the human evaluator, the two detection methods performed equally well. Towards the end of the dive the machine learning model detected divers on several occasions. However, most of these detections were not sequential so it is possible they were false positives, coincidental with divers in the water, vice actual detections.

It is significant that there was a large disparity between the dives concurrent with construction and the lower background noise dives. The dives during construction, 19 and 27 February, had a pre-filtering accuracy of 83.1% compared to the human operator. This rose to 90.4% after filtering. The 30 January and 6 February dives, which had limited background noise, had a pre-filtering accuracy of 98.2% compared to the human evaluator. This changed to 99.5% after processing was applied.

Overall sequential processing proved to be beneficial in most circumstances. This was because it filtered non-consecutive false positives with a minimal cost to integration time, improving the overall performance of the machine learning model. Sequential processing with a medium confidence detection threshold successfully removed 61.5% of false positives and 64.5% of missed detections, bringing the systems accuracy from 93.0% prior to filtering to 97.1% after filtering. This demonstrates that sequential processing is beneficial under the majority of circumstances.

4.6 Acoustic Arrival Path Determination

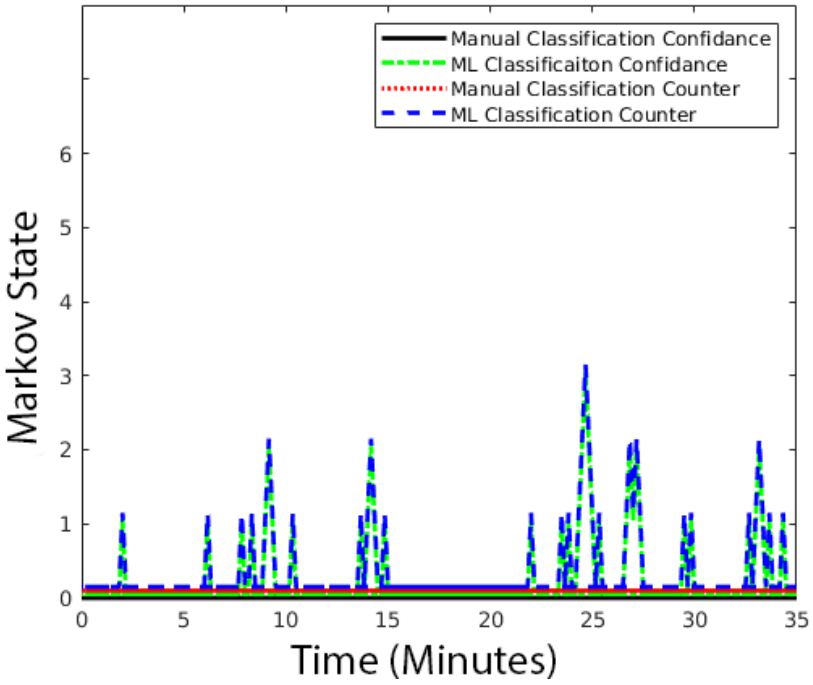


Figure 4-16: Sequential Data Processing, Divers not Present, 27 February 2019.

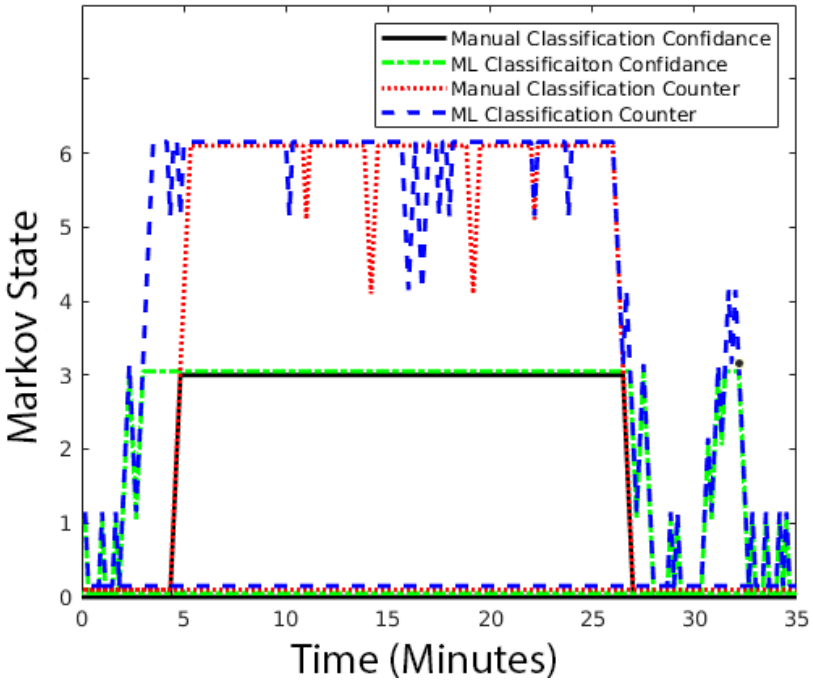


Figure 4-17: Sequential Data Processing, Divers Present, 27 February 2019.

Acoustic modeling was used to determine the arrival path between the divers and the hydrophone. The analysis shows that multi-path propagation was present. The use of 3 arrival paths out performed direct path only propagation by 26.4%. Using 9 arrival paths in place of 3 only improved performance by an additional 2.5%, indicating more than three arrival paths were present; however the additional arrival paths only marginally improved performance. The dominant means of acoustic arrival were direct path and bottom bounce. This was driven by the fact the divers and the hydrophone were located close to the bottom, the water was relatively deep, and the dominant source of acoustic loss was spherical spreading. Any arrival path that included a surface bounce had a significantly longer path length and therefore more loss. With divers at a range of 3.05 m the surface bounce path contributed approximately 1/8 of the acoustic pressure of bottom bounce. When divers were at a range of 15.24 m the contribution of the surface bounce lowered to 1/12 that of the bottom bounce. In shallower water surface bounce propagation paths would have been more significant and therefore likely increased detection range.

Table 4.5: Acoustic Model Error as a Function of Arrival Paths.

Number of Arrival Paths	1	3	5	7	9
Mean Squared Error (μPa^2)	$1.9502*10^6$	$1.5429*10^6$	$1.5411*10^6$	$1.5393*10^6$	$1.5391*10^6$

Table 4.6: Predicted Acoustic Pressure as a Function of Diver Range and Number of Arrival Paths.

Distance	3.05m (10ft)	6.1m (20ft)	9.14m (30ft)	12.19m (40ft)	15.24m (50ft)
Measured Pressure (mPa)	23.665	11.564	4.771	4.808	3.828
1 Path Predicted Pressure (mPa)	24.16	8.854	6.021	5.03	4.571
3 Path Predicted Pressure (mPa)	24.133	9.201	6.045	4.898	4.359
5 Path Predicted Pressure (mPa)	24.133	9.203	6.046	4.898	4.356
7 Path Predicted Pressure (mPa)	24.133	9.204	6.046	4.897	4.353
9 Path Predicted Pressure (mPa)	24.131	9.205	6.047	4.898	4.354

Table 4.5 indicates the use of multiple arrival paths outperformed a single arrival path. It also shows that model performance rises as the number of arrival paths increase. However, figure 3-34, visually implies that multiple arrival paths only marginally outperformed the use of a single arrival path. Using mean squared error as the metric for the loss function, an improvement of 26.4% was achieved when changing from one to three arrival paths, while only an additional 2.5% improvement was gained by shifting from 3 to 9 arrival paths. It is important to note that the hydrophone used for this experiment was not calibrated and this likely contributed to some of the difference between actual and predicted pressure.

A total water depth of 21 m, as opposed to a shallower water depth, was likely the main reason that additional arrival paths only slightly improved model performance. Based on scattering and absorption coefficients equal to one and an attenuation loss of only 2 dB per kilometer, spherical spreading was the primary source of acoustic loss. Any arrival path that involved a surface bounce had a significantly longer path length, and resulted in more spreading loss.

For a diver 15.24 m (50 feet) from the hydrophone, the direct path length was 15.24 m, the bottom bounce path length was 15.37 m and the surface bounce path length was 43.335 m. As a result, the surface bounce only contributed approximately 1/12 of the acoustic pressure at the hydrophone compared to the bottom bounce path. Because the diver and hydrophone were both near the bottom, with a diver at a distance of 15.24 m (50 feet), the bottom bounce contribution to the received pressure at the hydrophone was approximately 98% of that of direct path. Using the same calculation for a diver 3.05 m (10 feet) from the hydrophone the surface bounce path contributed approximately 1/8 of the pressure of the bottom bounce path, and the bottom bounce path contributed approximately 70% of the pressure of direct path. All other arrival paths contributed less than the surface bounce arrival path.

The hydrophone use in this experiment was low cost and as previously motioned, not calibrated. As such the accuracy and precision of the hydrophone are questionable and should be considered when evaluating the conclusion drawn below. The fact that both of the coefficients for scattering and absorption were determined to be 1.0 suggested that even more than nine arrival paths were actually present, and because additional paths were not included, the model adjusted to give the arrival paths that involved at least a single bounce more weight than they should have received. This conclusion was based on the shape of

Table 4.7: Predicted Acoustic Pressure for Various Arrival Paths.

Path	Direct	Surface Bounce	Bottom Bounce	Double Bounce	Double Surface Bounce	Double Bottom Bounce	Four Bounce
Path Length (m) Diver at 3.05 m	3.048	40.7861	3.6456	62.166	83.3997	44.7759	85.3984
Path Length (m) Diver at 15.24 m	15.24	43.4335	15.3707	65.8439	84.7259	47.2001	86.694
Predicted Pressure Contribution (μPa) Diver at 3.05 m	$1.2197*10^4$	67.5262	$8.5246*10^3$	28.9236	15.9921	55.9769	15.2435
Predicted Pressure Contribution (μPa) Diver at 15.24 m	486.4975	59.5088	478.2464	25.7607	15.4906	50.3465	14.7886
Predicted Pressure / Direct Path Predicted Pressure Diver at 3.05 m	1	0.0055	0.6989	0.0024	0.0013	0.0046	0.0012
Predicted Pressure / Direct Path Predicted Pressure Diver at 15.24 m	1	0.1223	0.983	0.053	0.0318	0.1035	0.304

the measured and predicted pressure in figure 3-34 and the fact that no reflection will be without loss. Loss due to scattering and absorption was expected because the acoustic wave length of the signal, 8.57 cm, was not long compared to typical wave height, and the bottom was mostly soft mud, frequently covered with muscle and clam shells. The wave length is equal to speed of sound in water divided by the frequency. Using the mean frequency of 17.5 kHz, which accounts for filtering out the lowest 5 kHz of the signal and a speed of sound in water of 1500 m/s the wavelength was 8.57 cm.

$$\lambda = \frac{c}{f}$$

Table 4.7 shows the path lengths for the various arrival paths for a diver at a distance of both 3.05 m and 15.24 m (10 and 50 feet). It also shows the predicted pressure contribution in micro Pascals to the total pressure for each arrival path and the fraction of each predicted pressure compared to the direct path predicted pressure. It is noteworthy that both the double bounce and the four bounce arrival paths contribute twice to the total received pressure.

An additional source of error that may have contributed to the difference between measured and predicted pressure at the hydrophone was the way that the hydrophone was mounted. The hydrophone was rigidly mounted to a large metal frame placed on the seafloor.

It is possible that the frame both shielded the hydrophone from some of the acoustic energy, and that the frame transmitted acoustic energy that hit it directly to the hydrophone. Either of these phenomena would have added error to the pressure measured by the hydrophone.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 5

Conclusions and Recommended Future Work

This thesis makes an original contribution to the field of scuba diver detection and leads to several other conclusions that are discussed in this chapter. The original contribution was low cost passive sonar and a properly tuned convolutional neural network can be used to detect divers in noisy environments. Other notable conclusions are listed below.

- Sequential processing improved model performance.
- Multiple frequency bands increased the probability of diver detection.
- The model was able to detect divers from a single transient during brief lulls in background noise.
- Diver detection is a function of range and background noise.
- False positives were likely to occur during periods of heavy construction.
- Multi-path propagation existed between the divers and the hydrophone.

During this work several areas worthy of additional investigation were identified but not explored. These topics are recommended for future work and discussed in section 5.2.

5.1 Conclusions

This thesis examines the use of low cost passive sonar and machine learning for automated open circuit diver detection in a real world port environment. Twenty dives were conducted near the WHOI pier from October 2018 to February 2019. The acoustic energy present in the water was recorded with divers at known distances from the hydrophone. The recordings were converted into spectrograms that were used to train, tune, and validate a supervised machine learning model. Diver detection was conducted using an image processing approach to machine learning via a deep convolutional neural network. To the knowledge of the author this approach to open-circuit diver detection had not previously been attempted.

This thesis shows that deep convolutional neural networks are an effective method for open-circuit scuba diver detection with passive sonar. It further demonstrates that this method has several advantages over traditional automated diver detection. The advantages include the ability to detect divers from a single acoustic emission, no dependence on the number of divers, the ability to function under a wide variety of environmental conditions and the ability to perform in the presence of interfering contacts.

5.1.1 Sequential Data Processing

Sequential data processing improved the performance of the dive detection system. When applied, model performance increased from 93% to 97%, compared to a human operator. This suggests that sequential processing should normally be used in conjunction with the machine learning model. The limitation of this type of processing is that it increased integration time to 20 seconds; reducing the probability of diver detection during brief breaks in background noise. The result of improved classification accuracy by including previous data is likely transferable to to other systems that assess classification status on a recurring basis.

5.1.2 Multiple Frequency Bands to Improve Detection

Multiple frequency bands should be used to improve diver detection over a wide range of conditions. Two frequency bands, 8-15 kHz and 18-25 kHz, were used in this work. The choice to use these two bands is discussed in chapter 3. Divers were detectable only in the higher frequency band 8.6% of the time and only in the lower frequency band 7.4% of the

time. The lower frequency band was able to detect divers at a further range during low ambient noise conditions while the higher frequency band detected divers better during high background noise. This general result is likely transferable to other geographic locations and applications, but the optimal frequency and number of bands may vary, depending on the site specific conditions.

5.1.3 Diver Detection During Breaks in Noise

Deep convolutional neural networks are able to detect divers during brief lulls in high background noise. On 18 January 2019 there were three brief breaks in high background noise where divers were otherwise masked. During these breaks, four diver transients were detectable by a human. The machine learning model successfully detected divers in three of the four instances. Traditional automated diver detection systems, discussed in chapter 2, would have failed to detect divers during this time due to their requisite integration time. The model's performance demonstrates the flexibility of convolutional neural networks for diver detection. Convolutional neural networks only require a single data point to produce identify a object and therefore are likely advantageous for other classification problems.

5.1.4 Detection as a Function of Range and Noise

Diver detection is a function of both diver range and background noise. Lower background noise and shorter diver range result in higher probability of diver detection. This conclusion was drawn by examining the data with respect to both range and background noise. Dives were split into three categories of ambient noise and the average quantitative value assigned by the author during the manual review process was plotted with respect to range for each category. Figure 5-1 shows the detectibility of the diver signature is inversely proportional to both range and background noise level. This conclusion is likely transferable to other uses of passive sonar for detection, including Department of the Defense applications.

5.1.5 False Positives

The model produced false positives more frequently during periods of marine construction. Evaluation of spectrograms producing false positives indicated that model reported divers during abrupt changes from high background noise to low background noise with the duration of both low and high background noise in excess of one second. This was common during pile

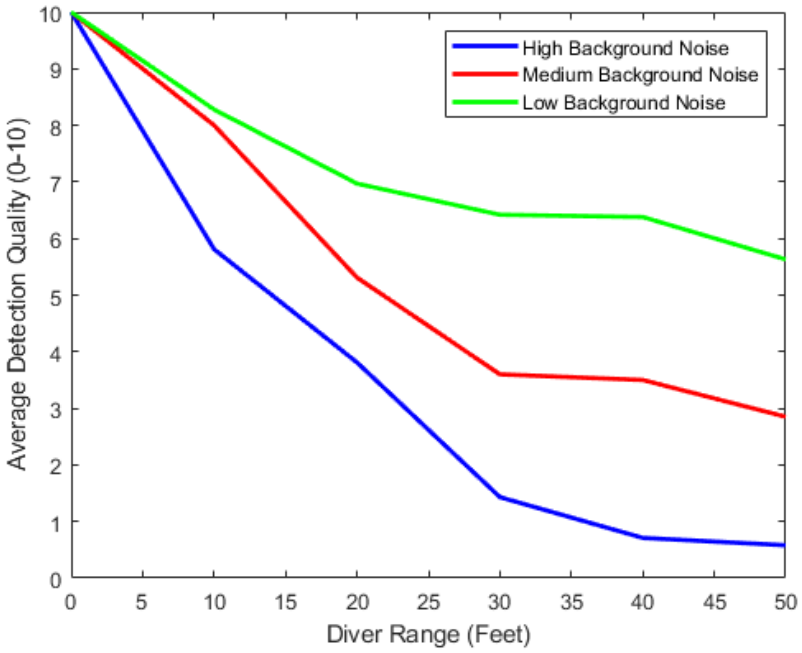


Figure 5-1: Diver Detection Quality as a Function of Range and Background Noise.

driving for nearby construction. The spectrograms producing false positives likely contained features closely resembling the features the model used for diver classification. It is possible that this type of false positive could be reduced by adding more construction data to training set prior to training the model. It would also be possible to produce a second machine learning model tuned specifically for diver detection during construction. Both options would be interesting follow on work and are discussed in section 5.2.

5.1.6 Propagation Path

Multi-path propagation existed between the diver and the hydrophone. Two arrival paths, direct path and bottom bounce, where responsible for the bulk of acoustic pressure at the hydrophone. Modeling suggested that there were at least nine arrival paths between the diver and the hydrophone; however, due to the relatively deep water depth and proximity of the divers to the bottom, any propagation path that involved a surface bounce contributed significantly less that direct path or bottom bounce propagation. This was a result of spreading being the dominate form of loss in the experiment. This suggests shallower water depth may result in longer detection ranges because the other arrival paths will be shorter

and therefore suffer less spreading loss. This result should be transferable to other detection problems with passive sonar.

5.1.7 Match Filtering

Match filtering between a known and suspected acoustic diver emission was not successful in aiding diver detection in this study. Cross correlation was noted between two diver emissions; however, when the reference signature was match filtered against broadband noise transients, the cross correlation was roughly 20 times larger than the two diver transients. This indicates that match filtering is not an optimal solution for diver detection in the presence of broadband transients. This conclusion was previously noted by Molbona and Zabarankin [33]. It is possible that there is a different method of match filtering or signal processing would have produced better results. This would be a topic worth exploring in future work.

5.2 Recommended Future Work

This thesis demonstrated that machine learning was a viable option for automated diver detection using low cost passive sonar. There is significant additional work to be done in the field of diver detection with machine learning including evaluating the model presented here in a different physical location, the use of other machine learning techniques, and the use of a hydrophone array for diver localization.

The model could readily be improved to perform better during construction activities. This could be done by adding more construction data to the non-diver training set. It is possible that doing this would make the model less sensitive to diver transients and therefore lower the model's performance. If training with more construction noise resulted in substantially degraded performance a second model could be developed for periods of construction. The normal model would be used in the absence of construction and the new model would be used during construction. It would be possible to shift models manually, automatically based on anticipated construction times, or automatically based on either the strength or duration of transients producing a classification of divers. If a transient indicated diver detection and was either significantly louder or of a duration different than expected for a diver the model would be shifted to the construction model. The model would be

shifted back to the normal model after a pre-specified time interval without any detections consistent with construction.

An important next step is to evaluate the machine learning model from this thesis in different locations. This would determine the degree of transfer learning that was possible and thus the cost of setting up a new diver detection system. Training data from several diverse locations could also be used to train and tune a new model in the hopes that using diverse training data would produce better results in different, previously unseen sites.

This thesis used spectrograms and image processing as the leaning medium; however, raw audio recorded by the hydrophone could also be used. A comparison of methods would be interesting; particularly if a combination of means were able to further reduce error rates.

A single passive hydrophone was used for this thesis. Using a single hydrophone provided omnidirectional results. This was a good first step; however, it is likely that performance could be improved by using an array of hydrophones. An array could potentially increase detection ranges and determine the bearing to the diver. This would be done by beam-forming the individual hydrophones on the array making the array more sensitive to bearings of lower background noise, as the noise from other directions would be suppressed. It would also provide the bearing from the array to the diver. Using two orthogonal arrays could potentially perform true diver localization. If each array could detect the diver and determine the bearing to the diver, the diver location could be determined by identifying the position where the two lines of bearing crossed.

Attempting to detect a closed circuit scuba diver using passive sonar and machine learning would be a topic of interest for follow on work. The acoustic signature of a closed-circuit scuba diver was not evaluated; however, it is likely significantly quieter than that of an open-circuit diver. This would make passive closed-circuit diver detection significantly more challenging, but if done successfully it could be combined with the work done in this thesis to make a system capable of detecting both open and closed-circuit divers.

Hari *et. al.* noted that diver inhalation transients were 1-1.3 seconds in duration [18]. Analysis of diver breathing transients in this thesis was consistent with Hari's findings; however, Hirsch and Bishop identified that the normal human inhalation duration is approximately two seconds [20]. It would be interesting to determine if the transient from diver inhalation was equal in length to diver inhalation or if it was a fixed length dependent on the first stage regulator. This could be determined by manually activating the regulation

for a specified period of time and evaluating the duration of the acoustic transient. This would be interesting follow-on work that could potentially be used for future diver detection systems.

An attempt was made to use match filtering between a reference diver transient and other diver transients. This work was moderately successful when unrelated broadband transients were not present. In the presence of broadband transients, the transients masked the cross correlation of the diver transients. With different signal processing or different techniques, match filtering of a reference diver signal and other diver signals may prove to be a viable means for diver detection in noisy environments. This analysis would be interesting follow on work.

As mentioned in chapter 1, the methods used in this thesis likely have applications beyond diver detection. The methods presented in chapter 3 are potentially useful for undersea warfare. An evaluation to determine if the work conducted in this thesis has defense applications would be prudent.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix A

Data Summary and Dive Information

Table A.1: Summary of Data Collected.

Dives	20
Days	15
Divers	7
Ships at WHOI Pier	3
Earliest Dive	0330
Latest Dive	1520
10 Second Spectrograms	5,474

Table A.2: Dive Information.

Date	Hour	Divers	Water Temp	Ship Present	Background Noise Level
10/5/2018	0900	Andrew, Ed	66F	Alucia Maru	Medium
10/5/2018	1000	Emmett, Fred	66F	Alucia Maru	Medium
10/5/2018	1100	Andrew, Ed	66F	Alucia Maru	Medium
10/5/2018	1100	Emmett, Fred	66F	Alucia Maru	Medium
10/5/2018	1200	Andrew, Ed	66F	Alucia Maru	Medium
10/19/2018	1000	Andrew, Ed, Emmett	62F	Atlantis	High
10/19/2018	1100	Andrew, Ed, Emmett	62F	Atlantis	Medium
10/26/2018	1500	Andrew, Ed, Emmett	55F	None	Low
10/31/2018	0300	Andrew, Ed, Emmett	50F	None	Low
10/31/2018	0500	Andrew, Ed, Emmett	50F	None	Low
11/30/2018	0900	Andrew, Ed, Joe	42F	Neil Armstrong	High
12/19/2018	0900	Andrew, Ed, Georgieo	39F	None	High
12/31/2018	Dive canceled Due to Excessive Construction Noise				
1/4/2019	0900	Andrew, Ed	39F	Neil Armstrong	High
1/8/2019	1200	Andrew, Ed, Joe	39F	Neil Armstrong	Medium
1/9/2019	0500	Andrew, Ed, Joe	39F	Neil Armstrong	Low
1/18/2019	0900	Andrew, Ed	37F	Neil Armstrong	High
1/30/2019	0500	Andrew, Ed	35F	Neil Armstrong	Low
2/6/2019	0500	Andrew, Ed, Joe	34F	Neil Armstrong	Low
2/19/2019	0900	Andrew, Ed, Kim	37F	Neil Armstrong	High
2/27/2019	0700	Andrew, Ed	37F	Neil Armstrong	Medium

Appendix B

Scuba Diving Procedure

Learned Diver Detection Data Collection Procedure:

Background:

The overall purpose of this experiment is to attempt to train a computer to do automated (open-circuit) diver detection based on passive sonar using machine learning techniques. This requires several data sets that record the divers' location (distance) with respect to time. This data will be used to determine the maximum range that the WHOI pier hydrophone can detect a diver, and will be used as a training or testing data set for automated diver detecting using machine learning. The questions that will be evaluated using machine learning include the presence or absence of divers, the range of the divers and the number of divers.

I thank you for your assistance in this data collection.

Required Equipment:

1. Tape measure on a reel (if not using pre-measured markers)
2. One or more sets of open circuit SCUBA equipment

   Note: Two or more divers are required to conduct this experiment, however only one needs to be using open circuit SCUBA equipment.

3. Underwater writing/recording device (slate or underwater notebook)
4. Underwater timing device

Procedure:

Note: ***Nothing in this procedure supersedes the principles of safe diving, or WHOI dive policies which must always take precedent.***

Note: *The hydrophone is located in 70 feet of water. The use of Nitrox is recommended as it maximizes the allowed bottom time for the divers.*

1. Prior to the dive review the data tables (on following pages) to ensure that you are aware of the required data for the experiment.
2. Fill out the pre-dive information which includes information about the divers and their equipment.
3. Enter the water and ensure that you and your partner are adequately situated and ready to proceed with the data collection.
4. Swim to the hydrophone and attach the tape measure to the base of the hydrophone using the snap-clip. This is only required if you are not using pre-measured markers along the lines along the pier.
5. Record the time that you are at the hydrophone (<5ft from the array). Remain at this location for 3-4 minutes.

6. Swim 10 feet farther from the array using the tape measure or pre-measured distance marker to note your distance. Record the time and your distance from the hydrophone. Remain at this location for 3-4 minutes.

    Note: Distances should be iterations of 10ft (ie 10ft, 20,ft, 30ft, 40ft, 50ft)

7. Repeat steps 6 as gas supply and dive limits permit.
8. Un-attach the tape measure from the hydrophone base (if required).
9. Complete the dive as normal.
10. Record post dive information and return the data sheets to Andrew Cole (Blake 201), coleam@mit.edu, 410 271-4545.

Pre-Dive Info

Date:_____ Number of Divers:_____

Weather (Sea State, Precipitation, Air Temp):_____

_____

Diver 1 Info:			
Name		Email	
Experience Level (Circle)		>100 Dives 20-100 Dives 0-20 Dives	
Open Circuit/Retreater			
First Stage Make/Model		Second Stage Make/Model	
Notes:			

Diver 2 Info:			
Name		Email	
Experience Level (Circle)		>100 Dives 20-100 Dives 0-20 Dives	
Open Circuit/Retreater			
First Stage Make/Model		Second Stage Make/Model	
Notes:			

*Additional Diver Information Blocks located at the end of the document.

Post Dive Information

Time Reference (EST/UTC)				
Time Entered Water		Time Exited Water		
Water Temp at Depth		Followed Procedure? Y/N		If no please annotate deviation in notes
Notes:				

Time / Distance Information:

Distance From Array (FT)	Diver Altitude (FT)	Time (Commenced) True time or time after dive start.	Duration (Min)

Time / Distance Information (Continued):

Distance From Array (FT)	Diver Altitude (FT)	Time (Commenced)	Duration (Min)

Additional Notes:

Diver 3 Info:			
Name		Email	
Experience Level (Circle)		>100 Dives 20-100 Dives 0-20 Dives	
Open Circuit/Retreater			
First Stage Make/Model		Second Stage Make/Model	
Notes:			

Diver 4 Info:			
Name		Email	
Experience Level (Circle)		>100 Dives 20-100 Dives 0-20 Dives	
Open Circuit/Retreater			
First Stage Make/Model		Second Stage Make/Model	
Notes:			

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix C

Manual Data Evaluation Form

Date		Hour		Record Freq	
DB Max		DB Min		Color Scheme	
Window		NFFT		Overlap	
Filter					

Minute	Dvr Dist	Detect Low	Detect High	Comments
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				

Minute	Dvr Dist	Detect Low	Detect High	Comments
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix D

Spectrogram Labeling Convention

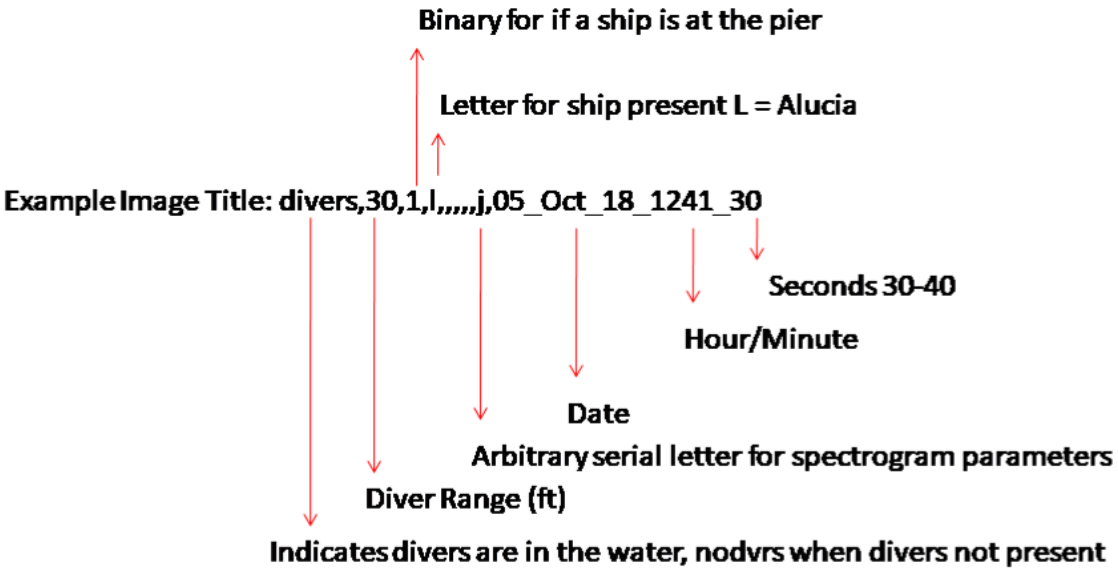


Figure D-1: Example Spectrogram Label.

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

[1] Keras: The Python Deep Learning library.

[2] MatPlotLib.org.

[3] SciPy.org.

[4] TensorFlow.org.

[5] MartÃŋn Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow : A System for Large-Scale Machine Learning. *Proceedings of the USENIX Symposium on Operating Systems Design and Implementation*, 12:1–21, 2016.

[6] Syed Anwar, Muhammad Majid, Adnan Qayyum, Muhammad Awais, Majdi Alnowami, and Khurram Khan. Medical image analysis using convolutional neural networks: A review. *Journal of Medical Systems*, 42:226, 10 2018.

[7] Dimitri P Bertsekas and John N Tsitsiklis. *Introduction to Probability*. Athena Scientific, Belmont, MA, second edition, 2008.

[8] Kevin Brinkmann and Jörg Hurka. A Multi-Target Tracking Approach for Narrowband Passive Sonar. *Narrowband Passive Sonar Tracking*, pages 8–10, 2010.

[9] Xiaoling Chen and Uf Tureli. Passive acoustic detection of divers using single hydrophone. *Conference Record - Asilomar Conference on Signals, Systems and Computers*, pages 554–558, 2006.

[10] Xiaoling Chen, Rensheng Wang, and Uf Tureli. Acoustic Detection of Divers Under Strong Interference. *IEEE*, (i), 2006.

[11] Kil Woo Chung, Hongbin Li, and Alexander Sutin. Frequency-Domain Multi-Band Matched-Filter Approach to Passive Diver Detection. *IEEE*, pages 1252–1256, 2007.

[12] Corinna Cortes and Vladimir Vapnik. Support-Vector Networks. *Machine Learning*, 20:273–297, 1995.

[13] Anna M. Crawford and D. Vance Crowe. Observations from demonstrations of several commercial diver detection sonar systems. *Oceans Conference Record (IEEE)*, pages 1–3, 2007.

[14] N. N. de Moura, J. M. de Seixas, and Ricardo Ramos. Passive Sonar Signal Detection and Classification Based on Independent Component Analysis. *Sonar Systems*, 2012.

[15] Dimitri M. Donskoy, Nikolay A. Sedunov, Alexander N. Sedunov, and Michael A. Tsionskiy. Variability of SCUBA diver's acoustic emission. *Proceedings of SPIE - The International Society for Optical Engineering*, (May):694515, 2008.

[16] J. R. OLMSTEAD ELFERS and T. H. MUSAC II : A Method for Modeling Passive Sonar Classification in a Multiple Target Environment A Method for Modeling Passive Sonar. (February), 1976.

[17] Kunihiko Fukushima. Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, 36:193–202, 1980.

[18] Vishnu N. Hari, Mandar Chitre, Yuen Min Too, and Venugopalan Pallayil. Robust passive diver detection in shallow ocean. *MTS/IEEE OCEANS 2015 - Genova: Discovering Sustainable Ocean Energy for a New World*, 2015.

[19] Geoffrey Hinton, Nitish Srivastava, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout : A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[20] J. A. Hirsch and B. Bishop. Human breathing patterns on mouthpiece or face mask during air, co2, or low o2. *Journal of Applied Physiology*, 53(5):1281–1290, 1982. PMID: 6816769.

[21] D. H. Hubel and T. N. Wiesel. RECEPTIVE FIELDS OF SINGLE NEURONES IN THE CAT ' S STRIATE CORTEX. *J. Physiol*, 148:574–591, 1959.

[22] D. H. Hubel and T. N. Wiesel. RECEPTIVE FIELDS, BINOCULAR INTERACTION AND FUNCTIONAL ARCHITECTURE IN THE CAT'S VISUAL CORTEX. *J. Physiol*, 160:106–154, 1962.

[23] A. T. Johansson, R. K. Lennartsson, E. Nolander, and S. Petrović. Improved passive acoustic detection of divers in harbor environments using pre-whitening. *MTS/IEEE Seattle, OCEANS 2010*, 2010.

[24] V. I. Korenbaum, S. V. Gorovoy, A. A. Tagiltsev, A. E. Kostiv, A. E. Borodin, I. A. Pochekutova, A. M. Vasilistov, A. C. Krupenkov, A. D. Shiryaev, and D. I. Vlasov. The possibility of passive acoustic monitoring of a Scuba Diver. *Doklady Earth Sciences*, 466(2):187–190, 2016.

[25] Alex Krizhevsky, Geoffrey E Hinton, and Ilya Sutskever. ImageNet Classification with Deep Convolutional Neural Networks. pages 1–9.

[26] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998.

[27] Yann Lecun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–444, 2015.

[28] R K Lennartsson, E Dalberg, L Persson, and S Petrovi. Passive Acoustic Detection and Classification of Divers in Harbor Environments. *IEEE Xplore*, (November 2009), 2009.

[29] Fei-Fei Li, Justin Johnson, and Serena Yeung. Training Neural Networks, Part 1, 2017.

[30] Kam W Lo and Brian G Ferguson. Diver Detection and Localization Using Passive Sonar. *Proceedings of Acoustics*, 8(November):1–8, 2012.

[31] Hannan Lohrasbipeydeh, Tom Dakin, T. Aaron Gulliver, and Claire De Grasse. Passive energy based acoustic signal analysis for diver detection. *2014 Oceans - St. John's*, pages 1–5, 2014.

[32] Madhav Mishra. Hands-On Introduction To Scikit-learn (sklearn), 2018.

[33] Anton Molyboha and Michael Zabarankin. Stochastic Optimization of Sensor Placement for Diver Detection. *Operations Research*, 60(2):292–312, 2012.

[34] Waseem Rawat and Zenghui Wang. Deep Convolutional Neural Networks for Image Classification : A Comprehensive Review. *Neural Computation*, 2449(29):2352–2449, 2017.

[35] Adrian Rosenbrock. A series of OpenCV convenience functions, 2015.

[36] Adrian Rosenbrock. *Deep Learning for Computer Vision with Python*. PyImageSearch, 1 edition, 2017.

[37] Adrian Rosenbrock. Deep learning, hydroponics, and medical marijuana, 2018.

[38] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, C V Jan, J Krause, and S Ma. ImageNet Large Scale Visual Recognition Challenge. 2015.

[39] Yutaka Sasaki. The truth of the f-measure. *Teach Tutor Mater*, 01 2007.

[40] Junaed Sattar and Gregory Dudek. A Vision-based Control and Interaction Framework for a Legged Underwater Robot.

[41] Pierre Sermanet, David Eigen, Xiang Zhang, Michael Mathieu, Rob Fergus, and Yann LeCunn. OverFeat : Integrated Recognition , Localization and Detection using Convolutional Networks. pages 1–16, 2014.

[42] Nabin S. Sharma, Alexander M. Yakubovskiy, and Matthew J. Zimmerman. SCUBA diver detection and classification in active and passive sonars - A unified approach. *2013 IEEE International Conference on Technologies for Homeland Security, HST 2013*, pages 189–194, 2013.

[43] Patrice Y Simard, Dave Steinkraus, John C Platt, One Microsoft Way, and Redmond Wa. Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. 2000.

[44] Hugh M. South, David C. Cronin, Samuel L. Gordon, and Timothy P. Magnani. Technologies for sonar processing. *Johns Hopkins APL Technical Digest (Applied Physics Laboratory)*, 19(4):459–468, 1998.

[45] Rustam Stolkin, Alexander Sutin, Sreeram Radhakrishnan, Michael Bruno, Brian Fullerton, Alexander Ekimov, and Michael Raftery. Feature based passive acoustic detection of underwater threats. *SPIE*, 6204:1–10, 2006.

[46] Zongxin Sun, Jiarong Zhang, Gang Qiao, Donghu Nie, Jialing Liao, and Songzuo Liu. Experimental Study on Target Characters of Divers. pages 1–5, 2013.

[47] Robert Urick. *Principles of Underwater Sound*. Peninsula Publishing, Westport, Ct, third edition, 1983.

[48] Petra Vidnerov and Roman Neruda. Evolving Keras Architectures for Sensor Data Analysis. In *Proceedings of the Federated Confrence on COmputer Science and Information Systems*, volume 11, pages 109–112, 2017.

[49] Alexander Waibel, Toshiyuki Hanazawa, Geoffrey E Hinton, Kiyohiro Shikano, and Kevin Lang. phoneme recognition using time-delay neural networks. *IEEE Transcriptions on Acoustics, Speach, and Signal Processing*, 37(3):328–339, 1989.

[50] Serdar Yegulap. What is TensorFlow? The machine learning library explained, 2018.

[51] Wu Zhao, Hong Chen, Longfeng Xiang, Xiaomei Xie, Min Chen, Zuli Zhao, and Qi Li. Passive Acoustic Detection of Diver Based on SVM. *IEEE International Confrence on Mechatronics and Automation*, 16:623–628, 2016.